

**Nástroj pro získávání cen z
prodejních portálů, sledování a
modelování vývoje trhu**

**Tool for Trade Portal Price Retrival,
and Market Trends Analysis**

Zadání bakalářské práce

Student:

Jaroslav Bouška

Studijní program:

B2647 Informační a komunikační technologie

Studijní obor:

2612R025 Informatika a výpočetní technika

Téma:

Nástroj pro získávání cen z prodejních portálů, sledování a modelování
vývoje trhu
Tool for Trade Portal Price Retrieval, and Market Trends Analysis

Zásady pro vypracování:

Cílem práce je sestavit nástroj pro získávání aktuálních cen z prodejních portálů, které nejsou zařazeny do produktových agregátorů. Následně pak nad získanými daty provádět analýzy aktuálního stavu trhu (cen, počtu prodejců, dostupnosti zboží) a dlouhodobé prognózy a modelování stávajících a budoucích trendů trhu.

1. Seznamte se s problematikou tvorby prodejních cen orientovaných na spotřebitele, nabídku, konkurenci atd.
2. Ve vhodně zvoleném prostředí proveďte analýzu, návrh a implementaci nástroje pro automatizované stahování cen produktů z prodejních portálů.
3. Definujte možné zdroje získávání informací o produktech, cenách a dostupnosti na internetu.
4. Sestavte databázi sledovaných produktů a vývoje jejich cen, na základě těchto informací odvoďte modely prodeje pro skupiny komodit.
5. Následně implementujte do sady funkcí, které budou popisovat stávající stav a pokusí se predikovat vývoj prodeje nebo celého trhu pro vybrané produkty.
6. Součástí implementace bude i nástroj, který bude porovnávat propagované ceny eshopu v rámci produktových vyhledávačů (Zboží.cz, Heureka) s reálnou nabídkou na stránkách. Bude automatizovaně upozorňovat na nesrovnalosti v cenách a dostupnosti zboží.
7. Výsledky vhodně vizualizujte v podobě tabulek, grafů a doporučení pro zadavatele.
8. Zhodnoťte dosažené výsledky a srovnajte je s existujícími postupy, následně pak uveďte další možnosti rozšíření.

Seznam doporučené odborné literatury:

Literatura podle pokynů vedoucího bakalářské práce.

Formální náležitosti a rozsah bakalářské práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí bakalářské práce: **Ing. Radoslav Fasuga, Ph.D.**

Datum zadání: 01.09.2013

Datum odevzdání: 07.05.2014



doc. Dr. Ing. Eduard Sojka
vedoucí katedry



prof. RNDr. Václav Snášel, CSc.
děkan fakulty

Souhlasím se zveřejněním této bakalářské práce dle požadavků čl. 26, odst. 9 *Studijního a zkušebního řádu pro studium v bakalářských programech VŠB-TU Ostrava*.

V Ostravě 20. dubna 2014

Bošk
.....

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Ostravě 20. dubna 2014

Bošk
.....

Rád bych na tomto místě poděkoval svému vedoucímu Ing. Radoslavu Fasugovi, Ph.D. za cenné rady, odborné vedení a připomínky při vzniku práce .

Abstrakt

Obsahem této bakalářské práce je problematika produktových agregátorů a nekalé činnosti provozované pomocí těchto nástrojů. Hlavním cílem práce je návrh a implementace nástroje pro získávání dat z internetových obchodů, které nejsou registrovány v produktových agregátorech. V první části práce se zabývám srovnáním čtyř tuzemských agregátorů produktů, kterými jsou Heuréka.cz, Zboží.cz, Google Nákupy a Srovnáváme.cz. Zaměřil jsem se na porovnání nabízených funkcí pro zákazníky i obchodníky a zdroje získávaných dat. V praktické části je provedena analýza, návrh implementace a vlastní implementace systému založená na technologiích PHP, MySQL a Java. Nástroj také obsahuje funkce pro zjištění nekalých činností v některých produktových agregátorech a funkce pro predikci vývoje ceny.

Klíčová slova: agregátor zboží, nekalé činnosti ve srovnávacích, parsování, predikce ceny, XML feed, Nette, PHP, Java, MySQL

Abstract

The main content of this work is the issue of product's aggregators and unfair business which are operated by these tools. The main goal of this work is design and implementation of tool for e-shop price retrieval, which aren't registered in products aggregators. In the first part of this work I deal a comparison of four domestic product's aggregators, which are Heuréka.cz, Zboží.cz, Google Shopping and Srovnáváme.cz. I compared functions for customers and businessman and source of the data obtained. The practical part of work contains the analysis, concept of implementation and implementation of the system using the technologies PHP, MySQL and Java. Tool also contains functions looking for unfair business in some product's aggregators and function for prediction of development price.

Keywords: product aggregator, unfair activities in the comparative, parsing, cost prediction, XML feed, Nette, PHP, Java, MySQL

Seznam použitých zkratek a symbolů

APEK	– Asociace pro elektronickou komerci
ASCII	– American Standard Code for Information Interchange
CPC	– Cost Per Click
CSS	– Cascading Style Sheets
CSV	– Comma Separated Values
DOM	– Document Object Model
EAN	– European article number
ELPOD	– Elektronické podnikání
FTP	– File Transfer Protocol
FTPS	– File Transfer Protocol Securedl
HTML	– HyperText Markup Language
HTTP	– HyperText Transfer Protocol
HTTPS	– HyperText Transfer Protocol Secured
ISBN	– International standard book number
JAVA	– objektově orientovaný programovací jazyk
JRE	– Java Runtime Environmentn
MVC	– Model-View-Controller
PHP	– Hypertext Preprocessor
PPC	– Pay Per Click
SQL	– Structured Query Language
SSL	– Secure Sockets Layer
SŘBD	– Systém řízení báze dat
URL	– Uniform Resource Locator
XML	– Extensible Markup Language

Obsah

1	Úvod	6
2	Problematika prodejních portálů	7
2.1	Historie obchodování na internetu	7
2.2	Produktový agregátor	7
2.3	Nekalé praktiky	9
3	Přehled produktových agregátorů	12
3.1	Heuréka.cz	12
3.2	Zboží.cz	15
3.3	Nákupy Google	17
3.4	Srovnáváme.cz	20
3.5	Porovnání agregátorů	21
4	Specifikace zadání	23
4.1	Získávání cen	23
4.2	Využití nástroje	23
5	Analýza	25
5.1	Datová analýza	25
5.2	Datová analýza	27
6	Návrh implementace	29
6.1	PHP a NETTE	29
6.2	JAVA	30
6.3	MySQL	30
7	Implementace	31
7.1	Webová aplikace	31
7.2	Aplikace pro získávání dat	36
7.3	Grafová komponenta	37
8	Predikce vývoje ceny	39
8.1	Zboží ovlivněno externími vlivy	39
8.2	Očekávaný vývoj ceny	39
8.3	Zboží s trendem růstu ceny	40
8.4	Sezonní zboží	40
9	Zhodnocení výsledků	43
9.1	Současný stav a testování	43
9.2	Vize do budoucna	43
10	Závěr	45

11 Reference	46
Přílohy	48
A Obsah CD	49

Seznam tabulek

1	Tabulka s přehledem funkcí agregátorů	22
2	Datový slovník	28
3	Výsledky testování pro import dat	44
4	Výsledky testování pro aktualizaci dat	44

Seznam obrázků

1	Obrat internetových obchodů v letech 2001 - 2012 [4]	8
2	Ukázka manipulace s klíčovými slovy [7]	11
3	Produktová stránka Heuréka.cz s oranžovým boxem	13
4	Produktová stránka Zboží.cz	16
5	Reklama Google AdWords s výsledky pro Nákupy Google	18
6	Ukázka využití rich snippets	24
7	ER diagram databázového schématu	26
8	Tvorba nové šablony	32
9	Přehled šablon pro vybraný internetový obchod	33
10	Přehled produktů v projektu	34
11	Detail produktu s cenami a grafem s vývojem cen	35
12	Ukázka vygenerovaného grafu pro 2.5 dimenze	37
13	Graf vývoje ceny pevného disku Western Digital	40
14	Graf vývoje ceny fotoaparátu Canon	40
15	Graf vývoje ceny pamětní zlaté mince [35]	41
16	Graf vývoje ceny zimních pneumatik [34]	41
17	Graf vývoje minimální ceny Samsungu řady Galaxy S [33]	41
18	Graf vývoje maximální ceny Samsungu řady Galaxy S [33]	42
19	Graf s porovnáním vývoje Samsungu řady Galaxy S [33]	42

Seznam výpisů zdrojového kódu

1 Úvod

Téma mé bakalářské práce se zabývá problematikou elektronické komerce, která dnes zahrnuje celou škálu metod obchodování na internetu. Přesněji se v práci zabývám produktovými agregátory, které nám umožňují přehledně procházet nabídku mnoha internetových obchodů. Pojem produktový agregátor vyjadřuje internetovou službu, která obsahuje informace o produktech z různých internetových obchodů. Nabízí jejich porovnání cen a technických specifikací. Dnes již také celá řada agregátorů slouží pro zákazníky i jako tzv. nákupní rádce, který pomáhá uživatelům s výběrem hledaného produktu, nabízí také recenze produktů od samotných uživatelů a informace o kvalitách internetových obchodů.

V teoretické části práce se zabývám přehledem existujících služeb produktových agregátorů na českém internetu. Uvádím zde jejich základní popis, nabízené funkce, které poskytuje pro obchodníky i nakupující zákazníky a metody přenosu informací o produktech z internetových obchodů do produktových agregátorů. Dále se zde zabývám nekalými praktikami ze strany prodejců pro zlepšení pozic při vyhledávání v agregátorech.

Praktická část práce obsahuje návrh a implementaci nástroje pro získávání cen z internetových obchodů, které nejsou zařazeny do produktových agregátorů. Na internetu je celá řada prosperujících či méně prosperujících obchodů. Registrace v produktových agregátorech s sebou nese jisté finanční náklady a ne každý provozovatel obchodu si jej může dovolit. Přesto by provozovatelé chtěli být informováni o cenách své konkurence, která v agregátoru nemusí být.

2 Problematika prodejních portálů

2.1 Historie obchodování na internetu

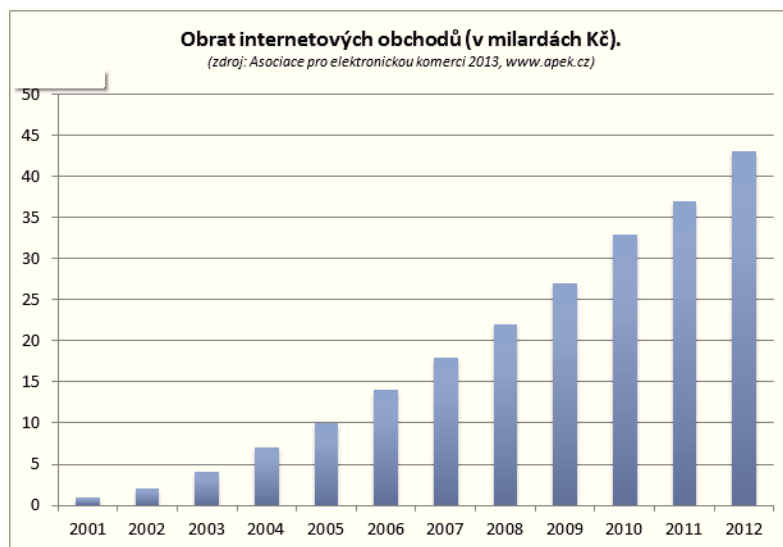
Obliba internetového nakupování v České republice neustále roste. Tuto oblibu dokazuje fakt rapidního nárůstu obrátu internetového prodeje, kdy v roce 2001 byl obrát jedna miliarda Kč a v roce 2012 více než 43 miliard Kč (Obrázek 1). Také vývoj můžeme pozorovat na nově vznikajících internetových obchodech, kde jich dnes je počet přesahuje 37 tisíc. Podle průzkumů měsíčně vzniká 800 internetových obchodů, přičemž jich 500 – 600 časem zanikne. Novým trendem v budování obchodů je specializace na jednotlivé produkty, kde se zvýšil počet o 16%, tento údaj vyplývá ze srovnávače cen Heuréka.cz.

Historie internetových obchodů sahá až do roku 1992, kdy se nejprve vyvíjeli v USA a sloužili především k distribuci audionahrávek. Typické internetové obchody s dnešní podobou začali vznikat v roce 1994, především díky vzniku SSL šifrování a tedy i možnosti plateb pomocí platebních karet přes internet. V témže roce vznikla společnost Amazon a aukční systém Ebay, kteří jsou dodnes hlavními obchodníky na trhu. Na českém trhu se začali objevovat internetové obchody od roku 1996, prvními z nich byla Vltava, která byla prvně knihkupectví, nebo třeba Alzasoft, dnes přejmenováno na Alza.cz. Teprve ale až po roce 2000 začínají čeští zákazníci vnímat internetové obchody jako bezpečné díky profesionálnímu přístupu online prodejců.[5, 1]

2.2 Produktový agregátor

Jak jsem již zmínil produktový agregátor, nebo také srovnávač zboží slouží především ke zlepšení orientace zákazníků mezi velkým množstvím nabízených produktů internetových obchodů. Ty se dělí na porovnávače zboží a nákupní rádce. Porovnávače zboží slouží zákazníkovi pro vyhledání jednotlivých produktů a porovnání cen různých internetových obchodů. Typickým představitelem je například Zboží.cz. Nákupní rádce se snaží zákazníkovi radit již při výběru zboží, porovnávat produkty různých výrobců a to buď na základě technických parametrů anebo uživatelských recenzí. Zde je typickým představitelem Heureka.cz. Zařazením obchodu do agregátoru se zvyšuje značně návštěvnost internetového obchodu, zvyšuje se i jeho potencionální zisk z prodeje. Mezi největší produktové agregátory na českém trhu patří Zboží.cz a Heureka.cz, kteří společně ovládají 84% trhu.[2]

Zařazení do produktového agregátoru je u většiny poskytovatelů zdarma, ale je zde zpoplatněno řazení produktů při vyhledávání, které vidí koncový uživatel. Princip tohoto financování je známý aukční model Pay Per Click (cena za proklik, zk. PPC) např. ze služby Google Adwords. U této formy reklamy platí obchodník pouze za proklik uživatele ze srovnávače zboží do jeho obchodu. Oproti jiným formám reklamy na internetu, kde se např. platí za umístění na webových stránkách nebo podle počtu zhlédnutí, je pro obchodníka velmi výhodná. Obchodník nastaví maximální cenu, kterou je ochoten zaplatit



Obrázek 1: Obrat internetových obchodů v letech 2001 - 2012 [4]

za proklik uživatele do jeho obchodu. Obchodník neplatí maximální nabídnutou cenu, platí pouze cenu, která je potřeba zaplatit pro udržení pozice.

Novinkou produktových agregátorů je služba vytvoření objednávky přímo v agregátoru. Není zde potřeba navštívit internetové prodejce. Tuto službu testuje Heureka.cz.

Všechny produktové agregátory, které budu zmiňovat, nenačítají informace o cenách a skladových zásobách zboží přímo ze stránek internetových obchodů, ale obchodníci data poskytují nejčastěji pomocí XML souboru umístěného na svých webových stránkách. XML soubor, nebo také zvaný XML Feed, obsahuje veškeré informace o nabízených produktech. Každý agregátor požaduje vyplnění povinných elementů a případně doporučených elementů, které se kladně projeví při konečném řazení produktů. Povinné elementy nejčastěji obsahují informace o názvu produktu, názvu produktu s přídavky (např. obal zdarma), název výrobce, popis výrobku, konečná cena s DPH, zda je produkt skladem, datum doručení, EAN kód, ISBN kód, URL adresa produktu na internetovém obchodě prodejce, URL adresa obrázku produktu, unikátní identifikační kód v daném internetovém obchodě a zda se jedná o nový nebo použitý produkt. Mezi doporučené elementy patří např. URL adresa s videorecenzí produktu, nastavení ceny za proklik, jednotlivé parametry produktu, zařazení do kategorií nebo různé varianty výrobku.

Protože srovnávače cen pracují s obrovským množstvím produktů různého typu je třeba je správně spárovat. Párování produktů probíhá automaticky, nejčastěji podle názvu produktu, výrobce, EAN či ISBN kódu, kódu výrobce, kategorie, případně popisu produktu. Každé kritérium musí být splněno z určité procentuální části, pokud není, nedojde ke správnému párování položek. V případě párování se nejprve zjistí, zda pá-

rovaný produkt má EAN nebo ISBN kód a zařadí ho mezi ostatní položky podle tohoto kódu. Pokud ale produkt nebyl zařazen, začne párování podle textu. Nejprve podle názvu a výrobce, pokud zde není nalezena shoda, pokusí se systém párovat podle popisu produktu. Popisy produktů mívají prodejci velmi často stejné nebo podobné, protože čerpají ze stejných zdrojů. Některé agregátory při párování zohledňují i cenu s určitým rozsahem aby nedošlo ke špatnému spárování nesouvisejících položek nebo neoriginálních položek s originálními. Pokud agregátor najde více stejných produktů z rozdílných obchodů, tyto produkty spáruje a spárované produkty poté zasadí do kategorií. Když pro hledaný produkt agregátor nenajde produkty ke spárování, nebude zařazen do kategorií a pro koncového uživatele bude přístupný pouze pomocí fulltextového vyhledávání.

2.3 Nekalé praktiky

Někteří prodejci pro získání širšího spektra zákazníků informace v XML souborech záměrně upravují, pro nalákání do svých internetových obchodů. Nejčastěji se jedná o upravení nabízené ceny, skladových zásob zboží nebo manipulaci s klíčovými slovy v názvech poskytovaného produktu. Takovéto jednání je však v rozporu s pravidly, které se zavázali dodržovat při registraci v produktovém agregátoru. Tyto nekalé praktiky však poskytovatelé agregátoru mohou těžko dokázat, neboť počet agregovaných položek se pohybuje v řádech milionů. Porušení pravidel agregátoru je nejčastěji potrestáno dočasným vyřazením obchodníka z vyhledávání nebo zablokování přístupu, tzv. BAN na určitou dobu.







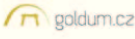



Hlavní důvod těchto nekalých praktik je zvýšení návštěvností internetového obchodu daného prodejce. Zákazníka již po vstupu na tento obchod může zaujmout jiná nabídka prodejce. Také ale toto jednání může mít pro prodejce z pohledu zákazníka negativní odezvu a již zde neuskuteční případný budoucí nákup. Nekalé praktiky internetových prodejců a klamání zákazníků:

- Upravení cen
 - prodejce upraví cenu produktu v datovém zdroji a poskytovat jej pro načtení do produktového agregátoru
 - cílem je zaujmout prohlížejího zákazníka na agregátoru a přivést je do internetového obchodu prodejce
- Dostupnost zboží
 - prodejce upraví informace o dostupnosti produktu
 - prodejce nabízí produkt, který ve skutečnosti nemá na skladě a neví datum jeho naskladnění
 - opět se snaží přivést zákazníky do obchodu případně nakoupit zboží
 - například obchodník nabízí zboží již před oficiálním prodejem a tvrdí že už produkt má skladem, ale ve skutečnosti pouze klame zákazníky

- Konečná cena
 - prodejci sníží nabízenou cenu o některou položku, např. o recyklační poplatek
 - v agregátorech musí být uvedena konečná cena produktu a to včetně DPH, recyklačních, autorských a dalších poplatků
 - správci agregátoru sice tyto rozdílné ceny kontrolují, avšak s malou četností
- Cena za jednotku
 - prodejce v agregátoru uvádí cenu za jednotku, nikoliv za celé balení
 - například prodejce nabízí koberec a uvádí cenu za m², zákazník ale může nakoupit pouze celou roli koberce
- Chyby ve feedech – neaktuálnost
 - prodejci se často vymlouvají na neaktuálnost načtených dat v agregátoru
 - pokud je tento problém u jedné položky, doopravdy se může jednat o neaktuálnost načtených dat.
 - pokud však tento problém je u většího počtu položek, nebo se jedná o dlouhodobý problém, jde ze strany prodejců bezpochybně o klamání zákazníka
- Manipulace s klíčovými slovy
 - srovnávače zboží používají fulltextové vyhledávání, prodejci zde pro získání více zákazníků upravují klíčová slova použitá v názvech produktů
 - prodejce nenabízí produkt, ale slova, která jej charakterizují, také uvádí u jiných více či méně podobných produktů
 - ukázka na obrázku 2, kde prodejce zlaťáky.cz manipuluje s klíčovými slovy pro spárování i s jinými produkty
 - v klasických fulltextových vyhledávačích jako je například Google je tato praktika penalizována, ale u produktových agregátorů se na tuto praktiku neklade takový důraz

Pokud je zákazník obětí těchto nekalých praktik, může se obrátit na správce agregátoru a nahlásit problém. Většinou však upozornění proběhne bez povšimnutí, protože produktový agregátor má z nabízených položek internetových obchodů zisk. Proto obchodníka nevyřadí z agregátoru natrvalo, ale pouze na určitou dobu.

Takovéto jednání bývá pouze začátkem nekalých úmyslů prodejců. Jak vyplývá z výroční zprávy České obchodní inspekce (ČOI), která za rok 2013 zkontrolovala 1296 [6] internetových obchodů a u 80% z nich zjistila porušení povinností. Nejčastěji se jednalo o již zmíněné nabízené produkty, které prodejce neměl skladem, nebo i dokonce neplánoval zboží naskladnit i přes zaplacení od zákazníka. U devítiset zjištěných případů byla uložena pokuta. Tato statistika je více než alarmující a ČOI vyzývá nakupující zákazníky k opatrnosti a k nákupu v již ověřených obchodech. Dále doporučuje nákup v obchodech, které vlastní certifikát APEK, ale takových obchodů je dnes registrováno pouze 311.

Nejlevnější nabídky Zlatý slitek Perth Mint 1 oz						
	 Ověřeno zákazníky ★★★★★ 135 recenzí	Investiční zlato - zlatý slitek 1oz Perth Mint	info v obchodu	26 447 Kč doprava od 75 Kč	Do obchodu	AuPortal
	★★★★★ 23 recenzí	Zlatý slitek The Perth Mint 1 Oz	do 3 dnů	26 768 Kč doprava od 89 Kč	Do obchodu	Zlateslity.com
	<u>Ohodnot' obchod</u>	Zlatý slitek 1 TROJSKÁ UNCE PERTH MINT (Austrálie)	info v obchodu	26 855 Kč doprava od 250 Kč	Do obchodu	BESSERGOLD.cz
	★★★★★ 29 recenzí	Investiční zlato The Perth Mint zlatý slitek 1oz	do týdne	27 039 Kč doprava od 100 Kč	Do obchodu	Zlatostribro.com
	★★★★★ 22 recenzí	zlatý slitek 1 oz Perth Mint	do 3 dnů	27 135 Kč doprava od 120 Kč	Do obchodu	Investiční zlato Pra
	 Ověřeno zákazníky ★★★★★ 53 recenzí	ZLATÝ SLITEK 1 OZ Perth Mint	do 3 dnů	27 354 Kč doprava zdarma	Do obchodu	GOLDUM
	 Ověřeno zákazníky ★★★★★ 611 recenzí	311g Perth Mint Investiční zlatý slitek	do 3 dnů	263 661 Kč doprava od 75 Kč	Do obchodu	zlataky.cz

Obrázek 2: Ukázka manipulace s klíčovými slovy [7]

3 Přehled produktových agregátorů

Přehled produktových agregátorů Tato kapitola se bude zabývat přehledem a srovnáním čtyřech nejpoužívanějších a nejznámějších produktových agregátorů v České republice. Jsou to Heureka.cz, Zboží.cz, Google Nákupy a Srovnáваме.cz. U těchto čtyř srovnávačů zboží provedu jejich základní popis a prozkoumám jejich funkce nabízené pro zákazníky, ale i funkce nabízené pro internetové obchody. Dále zde ukáži základní specifikace a datové soubory pro přeno dat z internetového obchodu do katalogu agregátoru. Porovnáám zde také hodnocení prodejců a jejich produktů při řazení a vyhledávání. Při řazení výsledků vyhledávání ovlivňuje často kromě ceny produktu, kvality prodejce, či recenze spotřebitelů zvýhodnění pozice a upřednostnění obchodu pomocí aukčního modelu ceny za proklik.

3.1 Heureka.cz

Je na českém internetu nejnavštěvovanější produktový agregátor zastřešující nabídku různých internetových prodejců a jejich nabízených produktů. Nabízí porovnávání produktů, jejich uživatelské recenze, ale také recenze internetových obchodů a jejich kvalitu. V současné době je v evidenci přes 20 tisíc obchodů. Oproti jiným srovnávačům zboží se liší tím, že neřadí výsledky vyhledávání produktů pouze podle ceny či předplacených zvýhodněných pozic, ale snaží se výsledky řadit podle ověřených prodejců, jejich kvalité a v neposlední řadě i podle subjektivních názorů samotných zákazníků. Také se od jiných agregátorů liší tím, že nabízí přímo prodej některých produktů a umožňuje nakoupit zboží v aukci přes portál Aukro.cz.

3.1.1 Nabízené služby

Heureka.cz nabízí uživatelům obrovské množství funkcí k porovnání a vybrání vhodného hledaného produktu. Uživatelé si mohou vybírat produkty podle zkušeností ostatních zákazníků. Zákazníci mohou ke každému produktu přidávat své hodnocení a zkušenosti. Heureka.cz zde připravila krátký dotazník jak je zákazník s produktem spokojen, může ho pomocí hvězdiček ohodnotit. V druhé části recenzent může zapsat do kolonek „Pro“, „Proti“ a „Shrnutí“ klady a zápory vybraného produktu a informovat o zkušenostech, které nabyl při jeho používání.

Další funkcí je ohodnocení obchodů, každému zákazníkovi po 10 dnech od objednávky přijde informační e-mail s dotazníkem ve kterém se může vyjádřit, zda byl spokojen či nespokojen s nákupem. V několika krátkých otázkách může zákazník informovat další zákazníky a majitele obchodu jak proběhl nákup, v jakém stavu dorazila objednávka, jestli již zákazník v tomto obchodě nakupoval, jak byl spokojen s přehledností obchodu a jiné. Obchod mohou hodnotit pouze zákazníci, kteří v něm nakoupili. Nelze tedy zhoršovat hodnocení obchodu např. cíleným konkurenčním bojem. Majitel obchodu tak získá zpětnou vazbu od zákazníků a zjistí, jak jsou spokojeni s nákupem, a na případné recenze může reagovat. Pro tuto funkci musí obchodník vložit do svého internetového obchodu



Obrázek 3: Produktová stránka Heureka.cz s oranžovým boxem

kód s unikátním identifikačním číslem. Dále musí dodat e-mail zákazníka a popřípadě název produktu, který je nepovinný. Po dokončení objednávky jsou tyto informace zaslány pomocí URL adresy na server Heureka.cz, kde budou uloženy. Když určitý počet zákazníků ohodnotí daný obchod kladně, získá tak certifikát **Ověřeno zákazníky**. Díky tomuto certifikátu, bude obchod řazen výš, oproti jiným obchodům, který tento certifikát nemají. Certifikát je zobrazen v katalogu prodejců, a majitel obchodu jej může umístit i přímo do internetového obchodu. Tato služba je pro obchodníky zcela zdarma a umožňuje tak vytvořit kvalitní obchod a dosáhnout na lepší pozice při řazení produktů ve výsledcích.

3.1.2 Řazení obchodů

Protože se Heureka.cz snaží doporučit rychlé, spolehlivé a doporučené obchody netřídí výsledky vyhledávání produktů pouze podle ceny, jak je to zvykem u jiných srovnávačů zboží, ale řadí je podle čtyř kritérií.

- Recenze obchodu - hodnocení obchodu na základě uživatelských recenzí po dokončení obchodu
- Dostupnost zboží - obchody, které mají daný produkt skladem, jsou řazeny výše. Kontroluje se pravdivost uváděných údajů, a pokud obchod případně záměrně zkreslí údaje, může být penalizován
- Cena zboží - další podstatné kritérium, cena musí být uvedena jako konečná, včetně všech recyklačních nebo autorských poplatků. Pokud obchod uvede záměrně odlišnou cenu, může být opět penalizován
- Cena za proklik uživatele - obchod si nastaví maximální cenu za návštěvníka, kterou je ochoten zaplatit. Poté jsou obchody řazeny podle nabídnuté ceny a dalších kri-

térií. Obchod neplatí maximální nastavenou cenu, ale pouze cenu jakou je potřeba zaplatit, aby se udržel na dané pozici v řazení. Cena za proklik musí být nastavená, aby byla stejná nebo vyšší než cena v dané kategorii. Cenu za proklik si může daný obchod nastavit v administraci, kde má přehled o zbývajícím nakoupeném kreditu, historii plateb a prokliků. Pokud obchod dostane certifikát „ověřeno zákazníky“ může nastavovat vyšší cenu za proklik oproti jiným obchodům a tím zvýšit šanci o umístění mezi TOP 3 do tzv. „oranžového boxu“ 3 . V tomto boxu mohou být obchody takové obchody, které mají daný produkt skladem. Avšak cena produktu již v tomto boxu nehraje roli.

Pro zjištění efektivnosti reklamy pomocí PPC Heuréka.cz poskytuje nástroj měření konverzí, kde se může obchodník informovat a vyhodnotit investice vložené do Heuréky. Poskytuje omezenou a plnou implementaci. U omezené implementace zobrazuje pouze konverzní poměr v procentech, tedy poměr objednávek vůči proklikům. U plné implementace se může obchodník informovat kolik procent lidí z Heuréka.cz v obchodě opravdu nakoupilo a jaká je návratnost investovaných prostředků. Pro obě statistiky musí obchodník vložit na konec objednávky vygenerovaný javaskriptový kód ze systému, který vloží do HTML stránky.

3.1.3 Zdroje dat

Informace o produktech z obchodů se stahuje z jejich XML feedů. Prodejce může poskytovat základní XML feed, který se aktualizuje každé dvě hodiny, nebo pokud vlastní certifikaci „Ověřeno zákazníky“ může poskytovat dostupností XML soubor. Tento soubor se aktualizuje každých deset minut a obsahuje pouze aktuální informace o dostupnosti zboží v internetovém obchodě nebo fyzicky skladem na kamenné pobočce, kde si jej může zákazník ihned vyzvednout. Pro přehled zde uvádím minimální doporučený obsah XML souboru.

- SHOP – kořenový element, ve kterém jsou umístěny všechny produkty
- SHOPITEM – element pro každý produkt, obsahující informace o konkrétním produktu
- ITEM_ID – unikátní označení produktu v daném internetovém obchodě, slouží také pro spárování s informacemi z dostupnostního XML
- PRODUCT – název produktu včetně přívlasků jako „nabíječka zdarma“
- PRODUCTNAME – název produktu bez přívlasků. Slouží pro správné spárování produktů. Název by měl obsahovat pouze výrobce, řadu produktů, produktové číslo, barvu atd.
- DESCRIPTION – popis produktu a jeho specifikace, slouží pro fulltextové vyhledávání
- URL – odkaz na stránku produktu v internetovém obchodě

- PRICE_VAT – konečná cena včetně DPH
- DELIVERY – obsahuje pod elementy s informacemi o možnostech a cenách dopravy
- DELIVERY_DATE – doba dopravy od přijetí platby či přijetí objednávky u dobírky

3.2 Zboží.cz

Produktový agregátor Zboží.cz je služba, kterou poskytuje společnost Seznam.cz. Je to jeden z nejstarších srovnávačů v České republice a díky společnosti Seznam.cz, která má na českém internetu dlouholetou tradici.

3.2.1 Nabízené služby

Stejně jako ostatní agregátory nabízí služba Zboží.cz 4 hodnocení produktů. Produkty mohou hodnotit pouze zaregistrovaní a přihlášení uživatelé pomocí účtu na Seznam.cz. Poté mohou ohodnotit produkt. Celkové hodnocení produktu je spočítáno jako průměr předchozích hodnocení a to formou hvězdiček v rozsahu 1 – 5. Po ohodnocení produktu mají uživatelé možnost připsat také textový komentář, jak jsou spokojeni s výrobkem. Přihlášený uživatel může produkt hodnotit pouze jednou, pokud produkt ohodnotí znovu, původní hodnocení se smaže a zůstane pouze nové.

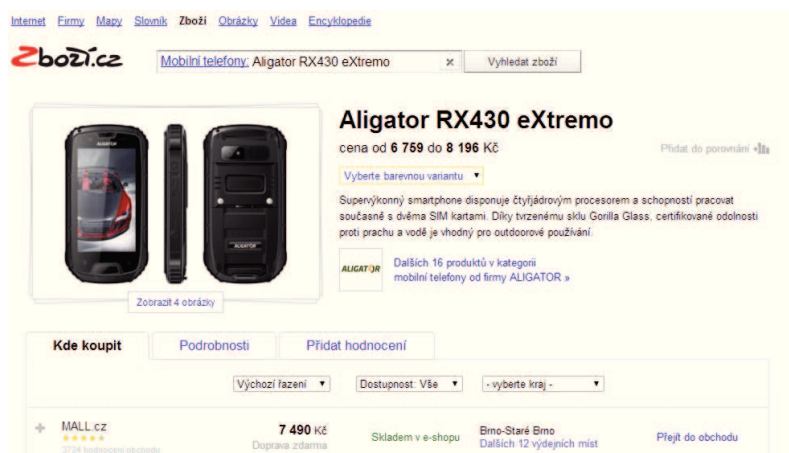
Hodnocení ochodů provádí také zákazníci. Hodnotit může zákazník, který provedl nákup v internetovém obchodě prostřednictvím Zboží.cz, nebo registrovaný zákazník pomocí služby Seznam.cz. Uživatelé, kteří provedli nákup, přijde po sedmi dnech email, který je vyzývá pro udělení hodnocení obchodu. Opět lze hodnotit pouze jednou, při opětovném hodnocení stejného obchodu dojde k přepsání původního. Pro recenzi obchodu složí formulář, kde uživatel pomocí formy hvězdiček vyjádří spokojenost s ochodem. Formulář se táže na pár základních otázek: Dodržení ceny a termínu dodání, komunikace eshopu, obsah a stav zásilky, popřípadě zkušenost s možností vrácení zboží a zkušenost s reklamací. Následně má uživatel možnost slovního hodnocení, kde může popsat zkušenosti s ochodem, spokojenost či dodací lhůta. Obchodník může pomocí administračního prostřední na tyto recenze reagovat, každé hodnocení může okomentovat jen jednou.

3.2.2 Řazení obchodů

Zboží.cz nabízí dva typy řazení produktů tzv. upřednostněné položky a neupřednostněné položky.

Neupřednostněné položky

- obchodník má přístup k základním statistikám - počty prokliků, CTR, počet zhlédnutí
- při řazení produktů ve fulltextovém vyhledávání jsou položky řazeny za upřednostněnými položkami



Obrázek 4: Produktová stránka Zboží.cz

- aktualizace XML feedu je 3x týdně
- cena za proklik je paušálně pro všechny produkty 0,2 Kč bez DPH

Upřednostněné položky

- obchodník má přístup ke statistikám, které jsou aktualizovány každých 24 hod.
- zobrazení položek na předních ve fulltextovém vyhledávání
- zobrazení proložek v produktových skupinách
- pokud má obchodník nastaveny u položek vyšší ceny za proklik než je 1 Kč bez DPH, jsou položky řazeny podle nastavené ceny a vynásobené kvalitou položek před položkami, které nemají nastavenou vyšší cenu za proklik
- aktualizace XML feedu je minimálně 1x za 24 hodin
- cena za proklik je paušálně pro všechny produkty 1 Kč bez DPH
- nastavení měsíčního rozpočtu

Pro samotné řazení produktů platí tyto dvě základní kritéria: maximální nabídnutá cena za proklik a kvalita daného obchodu. Kvalita obchodu závisí na uživatelském hodnocení, kvalitě dat v XML souboru a CTR (počet prokliků : počet zobrazení). Maximální cena za proklik se nastavuje přímo pro každý produkt zvlášť, nebo pro celé kategorie. Nastavit se dá pomocí administračního rozhraní, pomocí tagu MAX.CPC v XML souboru nebo pomocí importu CSV souboru. CSV soubor vygeneruje systém, obchodník si jej otevře, zapíše ceny pro jednotlivé produkty nebo celé kategorie a následně je importuje zpět do systému Zboží.cz. Maximální cena za proklik není odečítána, je odečtena pouze částka, která je potřebná pro dosažení dané nejlepší pozice. Částky za proklik jsou strhávány ze služby Seznam Peněženka, která musí obsahovat dostatek kreditu. Pokud

není co strhávat z nakoupeného kreditu, přestanou se zobrazovat nabídky produktů od prodejce. Lze také nastavit maximální měsíční rozpočet strhávaného kreditu, po kterém se přestanou zobrazovat nabídky produktů. Pro přehled úspěšnosti kampaní při řazení obchodů nabízí Zboží.cz statistiky. Pro aktivaci statistik musí prodejce na konec objednávky ve svém obchodu vložit HTML kód, který bude zajišťovat ukládání statistik. Tento kód po dokončení objednávky odešle informace o objednávce formou URL adresy na server Zboží.cz. Z těchto dat jsou následně generované výstupy ve formě tabulek a grafů.

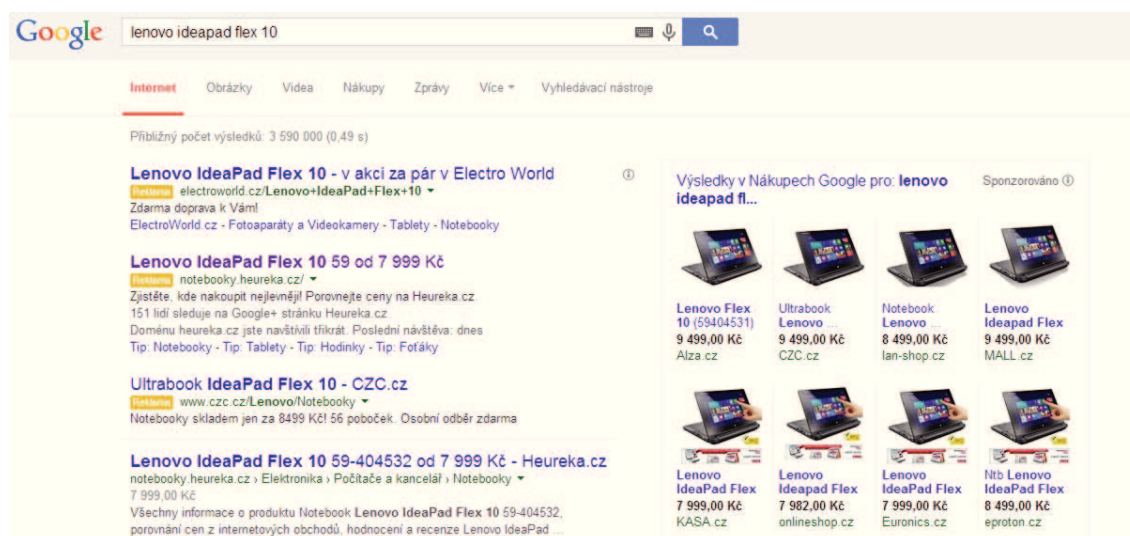
3.2.3 Zdroje dat

Data jsou získávána pomocí souboru XML, který je umístěn na webu prodejce. Při registraci zadá uživatel URL adresu s tímto souborem a data jsou pravidelně stahována. Soubor musí obsahovat základní nezbytné elementy s informacemi o produktu, pokud ale obsahuje i doporučené položky, objeví se to v pozitivně při řazení produktů. Seznam základních elementů:

- SHOP - kořenový element všech produktů
- SHOPITEM - element obsahující všechny informace o produktu. Obsahuje následující podelementy
- PRODUCT - název produktu, podle kterého lze jednoznačně identifikovat, může obsahovat informace typu "... + dárek zdarma"
- DESCRIPTION - popis produktu, který obsahuje výhradně informace o produktu a nesmí obsahovat reklamu
- URL - URL adresa s nabídkou daného výrobku
- PRICE_VAT - cena v Kč včetně DPH.
- DUES - všechny poplatky, které je potřeba zahrnout ke konečné ceně a nejsou zahrnuty u ceny s DPH. Nezahrnuje však dopravu a balné
- DELIVERY_DATE - doba vyřízení objednávky od přijetí po expedici zboží
- ITEM_TYPE - informace zda jde o nové nebo bazarové zboží

3.3 Nákupy Google

Nákupy Google je celosvětovou službou od společnosti Google umožňující vyhledávání a srovnávání cen jednotlivých prodejců na internetu. Kromě těchto služeb dále nabízí recenze jednotlivých produktů a možnost vytvoření si vlastního nákupního seznamu. Ačkoliv je to celosvětová služba, je kompletně přeložena do českého jazyka. Přímo na Nákupy Google nelze produkty nakupovat, pouze nám služba nabídne, kde hledaný produkt můžeme nakoupit.



Obrázek 5: Reklama Google AdWords s výsledky pro Nákupy Google

3.3.1 Nabízené služby

Hodnocení produktů je obdobné jako u jiných známých srovnávačů zboží. Recenze produktů lze provádět v detailu vyhledaného produktu pomocí uživatelského účtu z Google+. Je zde možné ohodnotit celkovou spokojenost produktu pomocí hvězdiček a slovního hodnocení. Zajímavou funkcí je odebírání recenzí z jiných serverů a to jak českých tak zahraničních. U recenzí ze zahraničních webů nabízí Nákupy originální znění, nebo strojový překlad recenze. Nákupy Google však nenabízí hodnocení obchodů, avšak recenze obchodů lze dohledat pomocí služby Google Adwords. Také pokud má internetový obchod kamennou prodejnu, lze zjistit spokojenost zákazníků pomocí Google Maps. Také zde lze přidávat recenze opět pomocí účtu Google+. Další funkcí je možnost nabízet bazarové či repasované zboží nebo zboží v aukcích. Zajímavou funkcí je propojení Nákupy Google se službou Google+ pro sdílení nákupního seznamu a komentování jednotlivých produktů.

3.3.2 Řazení obchodů

Nákupy Google je služba zdarma, avšak od 30. ledna 2013 nabízí komerční beta verzi, která umožňuje začlenění relevantní reklamy do výsledků vyhledávání Google. To znamená, že místo klasické placené reklamy s odkazy Google AdWords, může být uživateli zobrazena reklama přímo na prodejce hledaného produktu z Nákupy Google 5. Tato reklama je zpoplatněna podle ceny za proklik. Maximální cenu za proklik si může obchodník nastavit po propojení účtů Google AdWords a Google Merchant Center. Také zde si může nastavit maximální denní rozpočet, který je ochoten zaplatit za reklamní kampaň. Návratnost zaplacené reklamy si může prodávající zjistit pomocí komplexní služby Google AdWords, která nabízí mnoho dalších funkcí k úspěšnému podnikání.

3.3.3 Zdroje dat

Informace o produktech poskytuje prodejce a má na výběr ze dvou možností. První možnost je obdobná jako u českých služeb poskytováním XML souboru, který musí obsahovat určité povinné položky a doporučené položky. Druhou variantou je přenos dat pomocí textu ve formátu TXT odděleného tabulátory, kde se v prvním řádku uvedou názvy jednotlivých atributů. Na další řádky se již zadávají jednotlivé informace o produktech. U textových souborů je možno uvádět názvy atributů v češtině, nebo angličtině. U XML souboru musí být elementy v angličtině.

Přehled základních elementů:

- id, g:id - jedinečný identifikátor v rámci účtu obchodníka
- název položky, title - povinný název položky
- popis, description - popis položky
- kategorie produktu Google, g:google_product_category - obsahuje kategorii produktu dle taxonomie Google, např. Oblečení a doplňky - Oděvní doplňky - Sluneční brýle
- typ produktu, g:product_type - kategorie produktu podle obchodníka
- odkaz, link - URL adresa odkazující přímo na nabízený produkt
- URL obrázku, g:image_link - URL adresa obrázku dané položky
- URL adresa dalšího obrázku, g:additional_image_link
- stav, g:condition - stav položky, má přípustné pouze tři hodnoty, nové, použité a repasované
- dostupnost, g:availability - stav dostupnosti položky
- cena, g:price - koncová cena vč. DPH
- cena v akci, g:sale_price - inzerovaná cena položky v akci
- doba platnosti ceny v akci, g:sale_price_effective_date - časové období po kterou je položka v akci

Poskytované povinné atributy se liší v závislosti, do které kategorie nabízený produkt spadá. Ohraničující kořenové elementy v XML se také liší podle uvedené specifikace pro formát RSS nebo Atom. Zdroj dat můžeme nahrát přímo v administraci pomocí služby Merchant Center, nebo jej poskytnout na vlastním webovém či FTP serveru. Poté se zaregistruje URL adresa a případné přihlašovací údaje. V případě poskytnutí zdroje dat na vlastním webu, se budou data automaticky nahrávat a aktualizovat. Frekvence aktualizace lze nastavit v administraci. Počet aktualizací je omezen na denní limit počtu aktualizovaných produktů.

3.4 Srovnáváme.cz

Srovnáváme.cz je na třetí nejnavštěvovanější český produktový agregátor. Jako výše zmíněné služby nabízí porovnání jednotlivých produktů, podle ceny, vlastnosti produktů, zkušenosti zákazníků, dostupnosti zboží na prodejnách a nejbližších prodejen. Také samozřejmě nabízí porovnávání prodejců, jejich recenze a dodací lhůty. Pro prodejce, kteří chtějí na tomto agregátoru inzerovat je služba placená a to na dobu jednoho roku. Cena registrace závisí na počtu položek v XML souboru. Cena služby se také odvíjí od množství poskytnutých finančních prostředků pro PPC reklamu, která je využita v systému řazení produktů. Výše ceny prokliků je závislá na kategorii produktů a lze je navyšovat a dostat se tak na vyšší pozice ve vyhledávání. Při řazení produktů je také zohledněna kvalita nabízeného XML souboru a množství poskytnutých dat.

3.4.1 Nabízené služby

V detailu produktu uživatel může číst recenze a celkovou spokojenost s produktem. Také může samozřejmě produkt ohodnotit a to bez jakéhokoli přihlášení či autorizace. Je zde zaveden systém hodnocení pomocí hvězdiček pro celkové ohodnocení produktu a tři textová pole pro slovní hodnocení. První pole označené „PRO“ obsahuje informace v bodech o výhodách a přednostech produktu, popřípadě proč si jej zákazník zakoupil. Další pole je označeno „PROTI“ a zapisují se zde informace o záporech a nevýhodách produktu. Poslední pole „Celkové hodnocení“ umožňuje uživateli popsat vlastní zkušenosti s používáním produktu a další doporučení pro zákazníky. U hodnocení obchodů je obdobný princip, může jej recenzovat i nepřihlášený uživatel. To může, ale vést k úmyslnému poškození obchodu konkurencí, je zde však omezeno odeslání více recenzí z jedné IP adresy. Při hodnocení obchodu se systém ptá na tři základní otázky: „PRO, PROTI a Shrnutí“. Dále pomocí zaškrtačích políček se ptá na otázky „Komfort nakupování, Zákaznický servis a Včasné doručení“. Srovnáváme.cz nabízí certifikaci „Spolehlivý obchod“, pro kvalitní a ověřené prodejce. Certifikát lze získat, pokud je obchod držitelem certifikace APEK certifikát kvality nebo 3. testovacích anonymních nákupů a splnění hodnotících podmínek. Pro prodejce je zde služba měření konverzí, která informuje o výkonnosti kampaně a návratnost investic vložených do prokliků. Pro funkci měření konverze je potřeba vložit javascriptový kód na poslední stránku nákupního procesu v internetovém obchodě.

3.4.2 Zdroje dat

Data se načítají ze souboru XML, který poskytuje prodejce na svém webu a je možné použití formátu pro Heureka.cz, Zboží.cz nebo vlastní strukturu Srovnáváme.cz. Dle specifikace je uvedeno několik potřebných elementů, zajišťujících správné spárování produktů. Minimální struktura poskytnutého XML souboru.

- SHOP – kořenový element
- SHOPITEM– element, obsahující následující pod elementy s informacemi o produktu

- MANUFACTURER – název výrobce
- CATEGORYTEXT – celá cesta do kategorie
- EAN – kód výrobku
- ISBN – identifikační kód knih
- PRODUCTNAME – přesný název produktu
- DESCRIPTION – popis produktu
- URL – URL adresa produktu v internetovém obchodě
- IMGURL – URL adresa obrázku produktu
- ITEM_ID – unikátní a trvalé označení v rámci internetového obchodu
- PRICE_VAT - cena v Kč včetně DPH
- DELIVERY_DATE – počet dnů pro zajištění zboží

3.5 Porovnání agregátorů

Po pečlivém vyzkoušení a otestování funkcionalit jsem dospěl k závěru, že jako optimální řešení jak z pohledu zákazníků, tak z pohledu obchodníků je služba Heureka.cz. Snaží se doporučit zákazníkům spolehlivé obchody a při řazení výsledků zohledňuje spokojenost zákazníků s obchodem. Zajímavou funkcí je nákup zboží přímo v agregátoru, bez nutnosti návštěvy internetového obchodu. Tabulka ?? obsahuje srovnání zkoumaných agregátorů s výpisem jejich některých důležitých funkcí.

Zboží.cz je poměrně známá služba, fungující již řadu let. Z obvyčejné služby, která nabízela pouze vyhledávání zboží a výpis internetových obchodů kde mohou zákazníci zboží nakoupit, se stal plnohodnotný nákupní rádce, který pomáhá s výběrem vhodného produktu.

Google Nákupy jsou na českém internetu poměrně krátkou dobu a tak hledají stále své příznivce. Je to jednoduchá služba a oproti českým agregátorům nabízí i recenze zákazníků z celého světa. Jako za nevýhodu zde považuji málo informací o prodejci informace o spojenosti zákazníků s nákupem. Pro obchodníky však Google nabízí celou škálu nástrojů pro podnikání na internetu.

Služba Srovnáváme.cz se od jiných agregátorů zásadně neliší. Nabízí typické funkce, které od srovnávače zboží očekáváme. Prodejce nenutí generovat nový datový soubor XML pro import dat do systému, ale můžou použít již připravený např. pro službu Heureka.cz.

	Heuréka.cz	Zboží.cz	Google Nákupy	Srovnááme.cz
Recenze zboží	ano	ano	ano	ano
Bazarové zboží	ano	ano	ano	ne
Doplňkové zobrazení produktů	Zboží.Centrum.cz, Zboží.Atlas.cz, Nejlepsiceny.cz, SeznamZbozi.cz, SrovnaniCen.cz	vyhledávač Seznam	vyhledávač Google, reklama Google AdWords	ne
Recenze obchodů	ano, zákazníci, kteří nakoupili	ano, přihlášený uživatel ze Seznamu	ano (pomocí Google AdWords)	ano, kdokoli
Informace o obchodu	ano	ano	ano	ano
Řazení výsledků produktů	cena, dostupnost, specifikace, hodnocení	cena, název, specifikace	cena, název, specifikace, nové/použité	cena, název, specifikace
Řazení výsledků obchodů	cena, dostupnost, hodnocení, nejbližší obchod	cena, dostupnost, nejbližší obchod	cena, dostupnost, hodnocení, nejbližší obchod	cena, oblíbenost
Služby zdarma	zobrazování ve fulltextu na SrovnaniCen.cz	zařazení do neupřednostněných položek, aktualizace XML 3x týdně	ne	ne, požaduje minimální poplatek 900 Kč
Nástroj pro statistiky	ano	ano	Google Merchant Center a komplexní nástroj pro web Google AdWords	ano
Další služby	poradna, porovnání produktů, možnost nakoupit přímo na Heuréka.cz, hlídací pes	porovnávání produktů	sdílení nákupního seznamu pomocí Google+	porovnávání produktů

Tabulka 1: Tabulka s přehledem funkcí agregátorů

4 Specifikace zadání

4.1 Získávání cen

Mezi první způsob získávání cen o produktu je pomocí XML feedu. Je to nejjednodušší a nejrychlejší metoda, avšak nenabízí stoprocentní jistotu o konečné ceně, protože prodejce může tento soubor upravovat.

Další způsob jak získat aktuální informace o cenách je pomocí rich snippets 6[24]. neboli rozšiřující informace, které zavádí web Schema.org. Je to sdružení velkých společností jako Google nebo Bing a nabízí prodejcům umístit do speciálních tagů na webové stránce produktu informace o cenách a dostupnosti nabízeného zboží, popřípadě recenze zákazníků k produktu. Tyto informace pak samotný vyhledávač prochází a uživatelům nabízí přímo informace o produktu. Aktivace této služby není z pohledu prodejce nijak náročná, je potřeba lehce upravit zdrojový HTML kód a informace o produktu obalit do speciálních mikrodát, které nijak nenarušují vzhled nebo funkčnost obchodu. Z pohledu vyhledávání ceny je tato metoda poměrně přehledná a urychluje vyhledání ceny a dostupnosti. Bohužel skryté strukturované tagy dnes ještě nejsou příliš rozšířené.

Dalším způsobem je získávání ceny přímo z HTML stránky v obchodě prodejce. Je to nejpracnější, časově i technicky nejnáročnější metoda získání ceny a dostupnosti, avšak nejspolehlivější. Prodejce zde nemůže jakkoliv zasahovat do získávané ceny. Z technického hlediska je zde nejnáročnější vyhledat přesně a spolehlivě ceny či dostupnosti, protože různé internetové obchody používají různé platformy od různých výrobců, nebo někteří prodejci vytváří obchod sami bez jakýchkoliv dobrých programátorských návyků a vytváří zmatený kód. Nelze tedy přesně určit, kde přesně na webové stránce se nachází hledaná cena. Pro vytváření systému bylo použito vytvoření šablon individuálně pro každý internetový obchod. Řada internetových obchodů používá open source software nebo komplexní řešení obchodů od dodavatelských firem, které mají stejnou vnitřní strukturu zobrazovaného HTML kódu. Je tedy vytvořena šablona podle vyextrahovaného čistého textu z webové stránky. Tento text se vybírá podle ohraňujícího divu, který má jednoznačný atribut id nebo class. Pro správnou funkci šablony je třeba najít vhodný text před a za hledanou cenou či dostupností. Pokud není vhodný text před či za cenou, ale cena začíná na začátku nebo končí na konci vyextrahovaného textu, jsou zde použity vlajky, které oznamují, že před nebo za hledanou cenou není žádný text. Někdy je také potřeba určitou část textu přeskočit, ale pro jednoznačné nalezení hledané ceny musíme otestovat i text před textem přeskočeným. Po fázi přípravy vhodných šablon následuje fáze vyhledávání samotné ceny v textu.

4.2 Využití nástroje

I když většina internetových obchodů dnes je zaregistrována v některém produktovém agregátoru, stále velké množství obchodů zaregistrována není. Registrace v agregátoru sebou nese jisté finanční i technické náklady. V českých agregátorech jsou služby zpoplat-

Lacoste Essential, toaletní voda pro muže 125 ml | parfums.czwww.parfums.cz › Lacoste ▼

★★★★★ Hodnocení: 3 - 18 recenzí - 863,00 Kč - Na skladě

Toaletní voda Lacoste Essential je exkluzivní vůně pro muže, která ve vás vyvolá pocit absolutní svobody, která je také filosofií firmy ...

Lacoste Essential 75 ml | Alza.czwww.alza.cz › ... › Toaletní a kolínské vody › Toaletní vody ▼

★★★★★ Hodnocení: 4,3 - 4 recenze

Řekneme-li Lacoste, automaticky tím myslíme svobodu a radost. Právě tyto pocity v nás má evokovat i nová Lacoste Essential, z níž zaznívá apel: Udělejte si čas ...

Obrázek 6: Ukázka využití rich snippets

něny formou ceny PPC. Většinou sice poplatky začínají na korunových částkách za proklik na produkt, ale pokud má obchod zaregistrován např. XML feed s pěti sty položky, můžou se denní částky vyšplhat do řádů tisíců korun. To už ovšem nejsou zanedbatelné finanční položky, zvláště pro malé nebo začínající obchody. Dalším úskalím pro prezentaci nabízených produktů v agregátoru je generování XML feedu s informacemi. Je možné sice vytvářet XML soubor ručně, avšak je to velmi časově náročné. Prodejce zde může vytvořit řadu chyb a díky těmto chybám nemusí být XML soubor validně zpracován produktovým agregátorem, případně vůbec zařazen do výsledků vyhledávání. Velkou výhodou je používání tzv. opensource řešení internetových obchodů, které nabízí generování XML feedu ihned po instalaci bez žádných dalších doplňků. Takovým představitelem je třeba Prestashop.

Cílovou skupinou pro používání jsou začínající prodejci na internetu. Nástroj poskytuje kontrolu o cenách a skladových zásobách u konkurence. Nový prodejci tedy získají přehled o cenách na trhu. Dále je nástroj jako vhodný pomocník pro obchodníky, kteří sice jsou zaregistrovaní v produktových agregátorech, ale z důvodu velkého časového vytížení mohou zapomenout nakoupit kredit pro agregátor. Po vypotřebování nakoupeného kreditu dochází k deaktivaci a prodejce nebude řazen ve výsledcích vyhledávání. Ztrácí tedy viditelnost pro koncové zákazníky, ale také ztrácí přehled o cenách konkurence.

Dalším využitím produktu je získávání informací o nekalých aktivitách prováděných v produktových agregátorech. Nabízí porovnání cen Hereky či Zboží.cz s reálnými informacemi o cenách v internetových obchodech.

5 Analýza

5.1 Datová analýza

5.1.1 Lineární zápis typů entit

Typ entity, primární klíč, cizí klíč

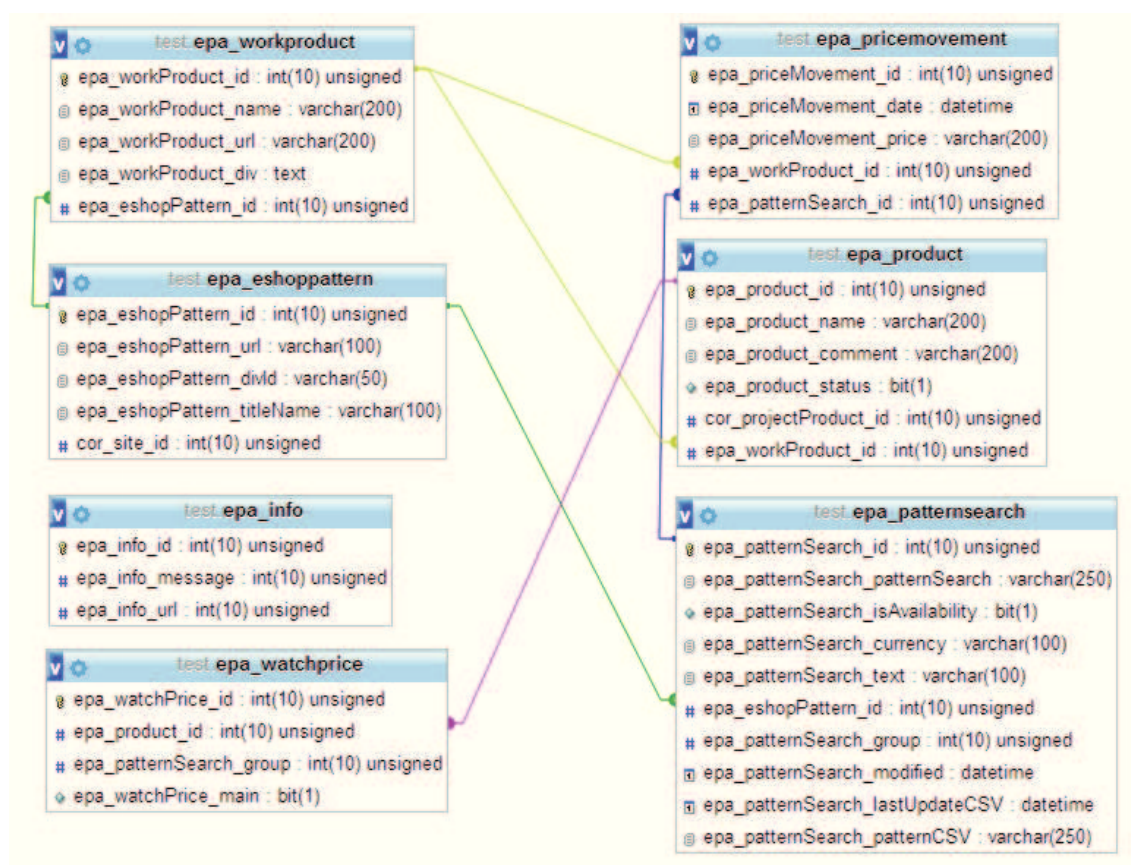
- **epa_product**(epa_product_id, epa_product_name, epa_product_comment, epa_product_status, *cor_projectProduct_id*, *epa_workProduct_id*)
- **epa_workProduct**(epa_workProduct_id, epa_workProduct_name, epa_workProduct_url, epa_workProduct_div, *epa_eshopPattern_id*)
- **epa_watchPrice**(epa_watchPrice_id, epa_watchPrice_main, epa_watchPrice_group, *epa_product_id*)
- **epa_eshopPattern**(epa_eshopPattern_id, epa_eshopPattern_url, epa_eshopPattern_divId, epa_eshopPattern_titleName, *cor_site_id*)
- **epa_patternSearch**(epa_patternSearch_id, epa_patternSearch_patternSearch, epa_patternSearch_isAvail, epa_patternSearch_currency, epa_patternSearch_text, *epa_eshopPattern_id*, epa_patternSearch_group, epa_patternSearch_modified, epa_patternSearch_lastUpdateCSV, epa_patternSearch_patternCSV)
- **epa_priceMovement**(epa_priceMovement_id, epa_priceMovement_price, epa_priceMovement_date, *epa_patternSearch_id*, *epa_workProduct_id*)
- **epa_info**(epa_info_id, epa_info_url, epa_info_message)

5.1.2 Lineární zápis seznamu vztahů

- has_eshopPatterns(epa_eshopPattern, epa_patternSearch) 1:N
- has_priceMovements(epa_workproduct, epa_priceMovement) 1:N
- has_patternSearchs(epa_patternSearch, epa_priceMovement) 1:N
- has_workProducts(epa_workProduct, epa_product) 1:N
- has_workProduct_eshopPatterns(epa_eshopPattern, epa_workProduct) 1:N
- has_watchPrices(epa_product, epa_watchPrice) 1:N
- has_site(epa_eshopPattern, cor_site) 1:1
- has_corProduct(cor_product, epa_product) 1:N

5.1.3 ER Diagram

ER digram datového modelu je uveden na obrázku 7



Obrázek 7: ER diagram databázového schématu

5.1.4 Datový slovník

Datový slovník je uveden v tabulce 2

5.2 Datová analýza

5.2.1 Funkce systému

- vytvoření šablon pro vyhledávání a parsování internetových obchodů
- přidání nových hlídaných produktů
- automatické stahování dat z obchodu
- analýza vývoje cen a vykreslení grafů

5.2.2 Vstupní data

- URL hlídaného produktu

5.2.3 Výstupy systému

- seznam hlídaných produktů – název, URL
- seznam pohybu cen k jednotlivým produktům – název, datum, měna
- informace o cenách pro dané produkty z Heureka.cz a Zboží.cz
- grafy s vývojem cen

5.2.4 Okolí systému

Administrátor

- může vytvářet a editovat šablony pro vyhledávání cen nebo dostupností
- může přidávat nové produkty, případně spravovat již existující

Registrovaný uživatel

- může přidávat nebo mazat hlídané produkty z konkurenčních obchodů
- může provádět analýzy historie vývoje ceny

epa_product			
název	typ	NULL	výchozí
epa_product_id	int(10)	ne	
epa_product_name	varchar(200)	ne	
epa_product_comment	varchar(200)	ano	
epa_product_status	bit(1)	ne	
cor_projectProduct_id	int(10)	ne	
epa_workProduct_id	int(10)	ne	
epa_workProduct			
epa_workProduct_id	int(10)	ne	
epa_workProduct_name	varchar(200)	ne	
epa_workProduct_url	varchar(200)	ne	
epa_workProduct_div	text	ne	
epa_eshopPattern_id	int(10)	ne	
epa_eshopPattern			
epa_eshopPattern_id	int(10)	ne	
epa_eshopPattern_url	varchar(200)	ne	
epa_eshopPattern_divId	varchar(50)	ano	
epa_eshopPattern_titleName	varchar(100)	ano	
cor_site_id	int(10)	ano	
epa_patternsearch			
epa_patternSearch_id	int(10)	ne	
epa_patternSearch_patternSearch	varchar(250)	ne	
epa_patternSearch_isAvailability	bit(1)	ne	
epa_patternSearch_currency	varchar(100)	ano	
epa_patternSearch_text	varchar(100)	ano	
epa_eshopPattern_id	int(10)	ne	
epa_patternSearch_group	int(10)	ne	
epa_patternSearch_modified	datetime	ne	
epa_patternSearch_lastUpdateCSV	datetime	ano	
epa_patternSearch_patternCSV	varchar(250)	ano	
epa_priceMovement			
epa_priceMovement_id	int(10)	ne	
epa_priceMovement_date	datetime	ne	
epa_priceMovement_price	varchar(200)	ne	
epa_workProduct_id	int(10)	ne	
epa_patternSearch_id	int(10)	ne	
epa_watchPrice			
epa_watchPrice_id	int(10)	ne	
epa_watchPrice_group	int(10)	ne	
epa_watchPrice_main	bit(1)	ne	
epa_product_id	int(10)	ne	
epa_info			
epa_info_id	int(10)	ne	
epa_info_message	varchar(200)	ne	
epa_info_url	bit(1)	ano	

Tabulka 2: Datový slovník

6 Návrh implementace

Pro různé části systému byly zvoleny různé platformy, které vyhovují daným specifickým požadavkům na řešení úkol. Pro webovou část aplikace byl zvolen skriptovací jazyk PHP, který se nejvíce hodí k tvorbě webových aplikací. Ostatní programovací jazyky nejsou přizpůsobeny pro tvorbu webových aplikací nebo vývoj v nich je značně zdlouhavý. Další výhodou PHP je velmi dobrá podpora komunikace s databází MySQL a navíc je multiplatformní. Pro ukládání dat byl použit MySQL server, který je pro tuto práci vhodný. Ke kontrole dat z internetových obchodů byl zvolen jazyk JAVA z důvodu vysokého výkonu oproti webovým technologiím. Také nabízí dobrou podporu k připojení k databázi pomocí knihoven JDBC.

6.1 PHP a NETTE

PHP[29] je serverový skriptovací jazyk, který umožňuje tvorbu dynamických webových stránek. Tvorba PHP skriptů je možná v obyčejném textovém editoru a nemusí se překládat. Pro jejich spuštění stačí webový server s podporou PHP. Syntaxe PHP je velmi podobná jazyku C. Má velmi strmou křivku učení, avšak programátoři v něm můžou snadno vytvořit chaotický kód. Proto jsem při vývoji použil frameworku Nette, který využívá návrhového vzoru Model-View-Presenter. Tento návrhový vzor vychází ze vzoru Model-View-Controller. Základním principem je oddělení kódu uživatelského rozhraní od kódu aplikační logiky, která je nazývána model. Uživatelské rozhraní je ještě rozděleno na vrstvu pro vizuální výstup (v tomto případě HTML kód) a na vrstvu obstarávající vstup uživatele. Vizuální výstup označujeme View a uživatelský vstup označujeme Presenter.

- Model - vrstva pracující s daty. Je kompletně oddělena od zbytku aplikace. Komunikuje pouze s Presenterem.
- View - front-end vrstva. Vykresluje požadovaná data pomocí šablon a zobrazuje je uživateli.
- Presenter (nebo Controller) - propojovací vrstva. Presenter propojuje Model a View. Zpracovává požadavky, dotazuje se Modelu na data a vrací je zpět do View.(citace z nette.org)

Díky tomuto striktnímu rozdělení může považovat webové aplikace za bezpečné. Pomocí konfiguračního souboru .htaccess (server Apache) zakážeme vstup přes http protokol do adresářů obsahující jednotlivé Modely nebo data pro připojení k databázi.

Nette Framework[27] je ryze český framework od autora Davida Grudla. I když se jedná o český projekt s poměrně malou komunitou, dokáže konkurovat i nejpoužívanějšímu PHP frameworku Zend. Hlavními přednostmi je použití objektově relačního mapování pro přístup k databázi a šablonovací systém Latte, který dokáže zabezpečit výstup před zranitelnostmi jako je XSS.

Pro potřebu parsování HTML stránek byla využita knihovna Simple HTML DOM Parser [25], která zajišťuje rychlé vytvoření DOM objektu z načteného HTML kódu případně URL adresy a umožňuje jeho efektivní procházení a vyhledávání tagů.

Pro potřeby vhodně vizualizovat průběhy vývoje cen byla použita JS knihovna Highcharts, která umožňuje generovat přehledné a interaktivní grafy. Pro plně automatizované vykreslování grafu byla vytvořena v PHP třída, která generuje JS kód grafu.

6.2 JAVA

Java je jedním z nejrozšířenějších programovacích jazyků na světě. Jeho hlavní výhodou je přenositelnost mezi různými platformami a umožňuje tvorbu od jednoduchých aplikací pro mobilní telefony přes webové aplikace až po rozsáhlé distribuované systémy. Je plně objektově orientovaný, robustní interpretovaný jazyk. Jednou z jeho výhod je velká množina knihoven, pod názvem JRE, která obsahuje např. funkce pro práci s databází (JDBC), více vláknového programování či práci s XML.

V jazyce JAVA bude realizována část aplikace, která bude zajišťovat automatickou kontrolu cen na prodejních portálech. V pravidelných časových intervalech bude stahovat HTML stránky ze serveru prodejce a kontrolovat dle nastavených šablon pro vyhledávání záznamy o pohybu cen a dostupnosti produktu.

Pro práci s HTML stránkami a jejich párování byla využita knihovna Jsoup [26], která zajišťuje rychlou manipulaci a extrahování dat pomocí DOM objektu. Pro výběr prvků používá známé selektory CSS.

6.3 MySQL

Ze všech databázových systémů, které jsou v současnosti k dispozici, jsem zvolil MySQL Server s datovým strojem InnoDB. Je pro tento projekt výhodný, protože podporuje standardy SQL, není placený a ve spojení s PHP je rychlý. Také jako PHP je podporovaný u většiny webhostingů a je multiplatformní. Jedním z velkých konkurentů MySQL je Oracle, který ovšem placený je.

7 Implementace

Samotná implementace nástroje pro získávání cen z prodejních portálů se dělí na dva základní bloky. Prvním blokem je samotné vytváření šablon pro jednotlivé internetové obchody, kde se pomocí průvodce vytvoří šablona pro vyhledání cen a dostupností na webové stránce produktu. Druhým blokem je již samotné vyhledávání cen podle vytvořených šablon. To probíhá automatizovaně v určitých časových úsecích pomocí aplikace v Java.

Při návrhu a samotné implementace systému, byl kladen velký důraz na jeho univerzálnost pro široké spektrum internetových obchodů. Při jeho vývoji byl testován na velkém počtu internetových obchodů.

7.1 Webová aplikace

7.1.1 Tvorba šablon a administrační část

Samotná tvorba šablon probíhá v administrační části systému, kde administrátor má přehled o produktech, pro které již šablony vytvořené jsou a o produktech, které si uživatel přidal, ale ještě pro ně není vytvořena vhodná šablona. Vytváření šablony probíhá pomocí krátkého průvodce, kde v prvním kroku je požadováno po správci vložení URL adresy s produktem a zapsání ceny. Webová stránka s produktem je stažena a načtena pomocí knihovny cURL. Dále pomocí třídy Simple HTML DOM Parser je vytvořen DOM objekt, který umožňuje efektivní procházení struktury načtené HTML stránky. Na stránce se vyhledají veškeré tagy div a je z nich načten prostý text. Protože různé e-shopy používají různé kódování souborů, jsou všechny texty převedeny na univerzální kódování UTF-8, které zahrnuje veškeré znaky včetně českých znaků. PHP neobsahuje žádnou přímou metodu na automatický převod textu mezi kódováními, je zde potřeba nejprve zjistit původní kódování textu. To se provádí opět prostřednictvím regulárních výrazů a podle zjištěného kódování je text převeden na UTF-8 za pomoci funkce iconv. Následně jsou z textu pomocí regulárních výrazů odstraněny veškeré zbytky HTML nepárových elementů, neviditelných znaků, nadbytečných mezer a HTML znakové entity jsou převedeny na viditelné znaky [30]. Dalším krokem je vytvoření regulárního výrazu, který umožní pomocí funkce preg_match vyhledat cenu v čistém textu. Toto je potřeba, protože opět každý e-shop má jinak naformátovanou cenu. Cena je vyhledávána v čistém textu, jelikož pokud by byla hledána v HTML kódu, systém by mohl nalézt nepravdivý údaj, např. ukázkovou cenu pro volání JS funkce na výpočet splátek při nákupu produktu.

Systém projde všechny nalezené tagy div, a pokud je v něm nalezena cena, je uložen div a jeho atribut class případně id do pole výsledků. Toto pole je předáno zpět správci a ten vybere vhodný div, který obsahuje hledanou cenu, případně text obsahující skladovou dostupnost produktu. V dalším kroku průvodce systém vytvoří informativní verzi šablony pro vyhledávání ceny. Pokud šablona neodpovídá požadavkům, může ji správce

Obrázek 8: Tvorba nové šablony

upravit pomocí posuvníku 8.

Vytvoření šablony pro hlídání dostupnosti probíhá také pomocí posuvníku, kde správce nejprve určí samotný text dostupnosti a pomocí druhého posuvníku určí okolní řetězec pro vyhledávání.

Dále zde může správce vytvořit vzor pro oříznutí názvu produktu, který se později při vkládání produktů načítá z HTML elementu title. Řada obchodů zde ale uvádí, o jaký web se jedná nebo zde uvádí jeho název, proto je vhodné uživateli vracet pouze název produktu, bez různých doplňujících informací. Správce také může vytvořit šablonu, která se bude pro danou cenu uživateli vypisovat.

Po upravení šablon a výpisů, jsou šablony uloženy do tabulek `epa_eshopPattern` a `epa_patternSearch`. Samotný řetězec pro vyhledávání ceny nebo dostupnosti musí mít optimální délku a obsahovat jednoznačné slova pro nalezení správné ceny. Na místo kde se vyskytuje cena, je vložen znak # a nástroj tedy ví jaký text má hledat před a za cenou. V některých případech se stává, že text neobsahuje žádný vhodný text před a za cenou. Proto jsem zde využil princip vlajek, které umožňují přeskakování textu. Jsou zde použity tři typy vlajek, start, end a next. Vlajka start je použita v případech, že cena se nachází ihned na začátku textu. Není tedy hledán žádný text před cenou, ale vyhledává se pouze text za cenou. Opačným případem je vlajka end která značí, že za nalezenou cenou již není žádný text. Vlajka next značí, že nástroj při hledání má přeskočit určitou část textu a pokračovat hledání až pomocí dalšího tetu v šabloně. Značka next je použita hlavně z důvodu, jelikož se na některých internetových obchodech objevují dva produkty na jedné stránce. Např. pokud vezmeme e-shop s mincemi obsahují na jedné stránce cenu proof a cenu běžné kvality mince.

Analýza eshopů - šablona pro web www.ceska-mince.cz

Přehled šablon Nová šablona Nová manuální šablona Přehled produktů Info

ID divu: Akce-zbytek-strany
 Vzor pro název: www.ceska-mince.cz - české mince jednoduše, snadno, přehledně.
 Web konkurence: www.ceska-mince.cz
[upravit](#)

Vzory pro vyhledávání:

Vzor	Měna	Šablona pro výpis	Je dostupnost	Skupina cen	Upravit	Smazat
ZM Proof: # Kč	CZK	Cena # Kč	ne	1	upravit	smazat
Do košíku + b.k.:#Kč	CZK	Cena # Kč	ne	2	upravit	smazat
Dostupnost zboží kód:# kód:		Dostupnost proof: #	ano	1	upravit	smazat
hodin kód:#Kredit		Dostupnost b.j.: #	ano	2	upravit	smazat

Vložit nový/upravit stávající:

Vzor pro vyhledávání:

Je vzor dostupnost: ☐

Měna:

Text pro výpis:

Skupina cen:

Obrázek 9: Přehled šablon pro vybraný internetový obchod

Ukázka výrazů pro šablony

- Správný tvar: Vaše cena vč. DPH: # Kč
- Správný tvar pro dostupnost: Stavby skladů: # Výrobce
- Špatný tvar: DPH: # Kč - při hledání ceny se nalezne cena včetně DPH, ale i cena bez DPH
- Použití značky start: –start– # Kč
- Použití značky end: Prodejní cena: # –end–
- Použití značky next: Provedení proof –next– ks # Kč, Provedení bk –next– ks # Kč

Nástroj tedy podporuje vyhledání více cen na jedné stránce. U některých typů e-shopů vlivem akčních nabídek nebo slev se mění struktura textu před nebo za cenou, nástroj proto umožňuje uložení více šablon. Šablony, které patří k sobě, např. Akční cena a Cena dnes mají udanou stejnou skupinu cen. To je využito při vyhledávání, nebo prezentaci dat. Pokud není nalezena cena podle prvního vzoru, nástroj se pokusí nalézt cenu podle některého dalšího vzoru ze skupiny.

Editace nebo tvorba nových šablon je možná v detailu šablony pro daný e-shop 9. Můžou se zde editovat veškeré prvky pro vyhledávání či zobrazování cen. Při vytváření nové šablony se může stát, že aplikace nenalezne cenu za pomoci regulárního výrazu. Administrátor je o tom informován a je přesměrován na vytvoření manuální šablony. Zde se již načítá webová stránka pouze pomocí URL adresy, je stažena a jsou získány všechny elementy div, které obsahují text. Opět se získá prostý text, který je převeden pomocí již výše zmíněných funkcí na UTF-8 a jsou odstraněny veškeré nežádoucí znaky. Ze všech textů je vytvořeno pole, které je vráceno správci pro vybrání vhodného textu.

Přehled podle produktů Přehled podle obchodů Přehled podle změny cen

Rychlé vyhledávání:

■ nejmenší cena
 ■ nižší jak průměrná cena
 ■ vyšší jak průměrná cena
 ■ nejvyšší cena

Produkt	Počet produktů	Všechny ceny	Maximální cena	Minimální cena	Průměrná cena
427. 5000 Kč Barokní most v Náměstí nad Oslavou, PROOF (15.05.2012)	7	17200	19900	17200	18243
429. 10000 Kč 800. výročí Zlatá bula sicilská, PROOF (26.09.2012)	6	27500	38700	37500	37767
430. 5000 Kč Negrelliho viadukt v Praze, BJ (13.11.2012)	6	18499	18900	14200	16686
431. 5000 Kč Negrelliho viadukt v Praze, PROOF (13.11.2012)	6	17999	19500	14850	16982
426. 5000 Kč Barokní most v Náměstí nad Oslavou, BJ (15.05.2012)	6	17799	18800	16800	18000
433. 5000 Kč Dřevěný most v Lenoře, PROOF (28.05.2013)	6	14899	17200	14899	16128
428. 10000 Kč 800. výročí Zlatá bula sicilská, BJ (26.09.2012)	6	39550	39550	37900	39183
436. 5000 Kč Železniční most v Žampachu, BJ (08.10.2013)	5	14850	15490	14645	14987
432. 5000 Kč Dřevěný most v Lenoře, BJ (28.05.2013)	5	14499	16800	14499	16188
423. 5000 Kč Gotický most v Písku, PROOF (25.05.2011)	4	34990	36000	33500	35072
417. 2500 Kč Kulturní památka větrný mlýn v Ruprechtově, PROOF (07.10.2009)	4	8450	9495	7800	8560
424. 5000 Kč Renesanční most ve Stribře, BJ (11.10.2011)	4	17999	19500	17999	18996
437. 5000 Kč Železniční most v Žampachu, PROOF (08.10.2013)	4	14899	15790	14890	15274
405. 2500 Kč Národní kulturní památka Klementinum v Praze - observatoř, PROOF (11.10.2006)	3	23300	31500	23300	28250
194. 200 Kč 100. výročí narození Bohumil Hrabal, BJ (26.03.2014)	3	449	479	425	458
425. 5000 Kč Renesanční most ve Stribře, PROOF (11.10.2011)	3		24950	21900	23662
195. 200 Kč 100. výročí narození Bohumil Hrabal, PROOF (26.03.2014)	3	549	599	525	568
403. 2500 Kč Národní kulturní památka papírna Velké Losiny, PROOF (10.05.2006)	3	66000	66800	61000	63700
435. 10000 Kč 1150. výročí Příchod věrozvěstů Konstantina a Metodie, PROOF (03.07.2013)	3	33365	33490	31980	32714
411. 2500 Kč Národní kulturní památka řetězový most ve Stádku, PROOF (14.05.2008)	2	9200	9450	9200	9325
189. 200 Kč 750. výročí Založení kláštera Zlatá koruna, PROOF (19.06.2013)	2	535	555	479	504
434. 10000 Kč 1150. výročí Příchod věrozvěstů Konstantina a Metodie, BJ (03.07.2013)	2	33950	33950	32990	33470

Obrázek 10: Přehled produktů v projektu

V dalším kroku správce vybere pomocí posuvníků cenu, šablonu pro cenu, případně dostupnost a šablonu pro dostupnost. Šablony jsou již dále uloženy stejným způsobem jako při předešlém vytváření šablon.

7.1.2 Přidávání a přehled produktů

Nástroj pro sledování a získávání cen z internetových obchodů je součástí webové aplikace, která zahrnuje komplexní nástroje pro podnikání na internetu. Produkty u kterých chceme hlídat pohyb cen, jsou spárovány s produkty, které již má uživatel zaveden v aplikaci u svého projektu.

Na obrázku 10 je ukázan přehled projektových produktů s již načtenými cenami. V jednotlivých sloupcích je uvedena minimální, maximální a průměrná cena. V barevném sloupci je načtena cena z produktů, které vlastní uživatel používající tuto aplikaci. Podle jednotlivých barev je vidět jestli má uživatel ve svém obchodě cenu nižší nebo vyšší než je průměrná, nebo jestli je jeho cena rovna maximální či minimální nabízené ceně z hlídaných produktů. Aplikace tedy umožňuje rychlý přehled a uživatel může reagovat na změnu cen konkurence. Prvním krokem pro přidání nového produktu je výběr projektu a produktu z již existujícího produktu. Uživatel přejde na detail produktu, kde má možnost přidat nové produkty pro hlídání. Pro přidání produktů, má dvě možnosti. První z nich je přidání produktu podle již zjištěné URL adresy. Druhou možností je vyhledat vhodné produkty na konkurenčních webech pomocí klíčových slov. Po výběru webové adresy konkurence a klíčového slova je vygenerován dotaz pro vyhledávač Google a je vrácen výsledek s nalezenými URL adresami. Uživatel si vybere adresu a zvolí přidání produktu.



Obrázek 11: Detail produktu s cenami a grafem s vývojem cen

Samotné přidávání produktů probíhá pro obě možnosti stejně. Nástroj ze zadané URL adresy zjistí doménu produktu. Podle zjištěné domény se pokusí vyhledat v databázi vhodnou šablonu pro daný internetový obchod, a jestliže je šablona nalezena, načtou se všechny šablony pro vyhledávání. Webová stránka s produktem je stažena a zpracována jako v případě vytváření šablon. Ve stažené produktové stránce je vyhledán div podle atributu id nebo class načtených z databáze. Text divu je opět zpracován a je vrácen prostý text v kódování UTF-8. V tomto textu již probíhá samotné vyhledávání ceny pomocí šablon. Nalezené ceny či informace o dostupnosti a název produktu jdou vráceny uživateli. Ten si může vybrat, které nalezené ceny bude chtít hlídat a vybere si jednu cenu jako hlavní, pro případnou prezentaci dat v tabulkách a grafech. Produkt je uložen do databázové tabulky `epa_product`, která ukládá základní údaje jako je vlastní název či komentář a stav o produktech pro každého uživatele. Tabulka napojena na tabulku `epa_workProduct`, kde se nacházejí skutečná data pro všechny produkty. Pro zabránění duplicity se data stahují a kontrolují pouze jednou, i když se produkt vyskytuje ve více projektech různých uživatelů. Jako unikátní klíč je zde použita URL adresa produktu.

Pod možností přidání produktu je tabulka obsahující hlídané produkty konkurenčních webů pro produkt projektu - viz obrázek 11. Je zde vypsána aktuální cena a dostupnost, případně cena načtena z Heuréka.cz či Zboží.cz. Pod tímto přehledem je znázorněn graf s vývojem hlídaných cen.

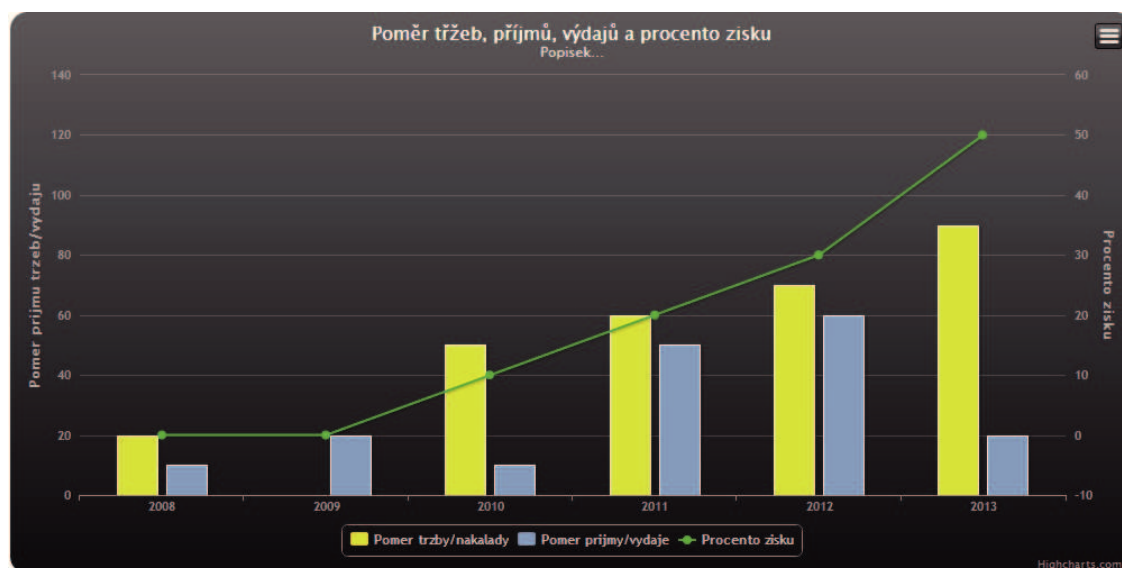
7.1.3 Vyhledávání ceny a dostupnosti

Pro samotné vyhledávání informací o cenách či dostupnosti jsem vytvořil jednoduchý avšak účinný algoritmus. Protože se data a text pro vyhledávání získávají z webových stránek, je zde velký problém s tzv. bílými znaky, které znemožňují vyhledávání pomocí výchozích funkcí PHP. Proto text šablony je rozdělen na jednotlivá slova a pomocí regulárních výrazů jsou odstraněny veškeré nežádoucí bílé znaky. Ze samotných slov je vytvořeno pole a začíná samotné vyhledávání. Nejprve se v textu nalezne první slovo ze šablony a je odstraněn předcházející text. Tak se pokračuje, dokud se nenarazí na znak #, který značí místo pro cenu. Za znakem # se pokračuje s vyhledáváním textu za cenou. Zde již není text ořezáván, nástroj si však pamatuje pozici prvního slova šablony za cenou. Pokud jsou nalezena veškerá slova šablony je odstraněn zbývající text za cenou. Pokud šablona obsahuje vlajky `-next-`, `-start-` nebo `-end-`, část textu přeskočí nebo vyhledává pouze před nebo za cenou. Pokud není nalezeno jedno slovo z výrazu, vyhledávání se okamžitě ukončí a pokračuje se vyhledáváním pomocí další šablony. To je provedeno, jelikož slova v šablony se již, mohly vyskytnout jinde na stránce, avšak další slovo již nemuselo být nalezeno. Vznikla by tedy nesmyslná vyhledaná cena obsahující velké množství textu. Výsledná vyhledaná cena je opět pomocí regulárních výrazů očištěna od přebytečných nežádoucích znaků. Pro hledání dostupnosti je aplikován stejný algoritmus.

7.2 Aplikace pro získávání dat

Pro velmi časově i technicky nákladnou aktualizaci dat o změnách cen jsem upustil od webové aplikace a vytvořil jsem pomocnou aplikaci v Java. Pro celkové urychlení běhu aplikace a zkrácení času aktualizace jsem vytvořil více vláknové zpracování dat. Časově nejnáročnější částí aplikace se projevilo stahování webových stránek. I když si aplikace stahuje pouze potřebný HTML kód o velikosti pár kB, je to nejpomalejší část celé aplikace. Pro stažení webové stránky produktu je použita knihovna Jsoup, která umožňuje i následné efektivní procházení divů s textem. Oproti PHP zde odpadá potřeba aplikování regulárních výrazů pro odstranění nebo převedení přebytečných a nežádoucích znaků jako jsou entity nebo bílé znaky. Knihovna také umožňuje automatickou konverzi do UTF-8. Pro získání dat z databáze a následnou práci s nimi jsem naimplementoval zjednodušené objektově relační mapování.

Aplikace po spuštění získá z databáze veškeré potřebná data o produktech, šablonách pro vyhledávání a o pohybu cen. Poté zkontroluje, zda mají všechny stažené šablony vytvořeny a aktuální CSV formát pro urychlení vyhledávání. Pokud nemají nebo byla šablona změněna, převede se šablona do CSV formátu, kde jako oddělovače mezi slovy slouží středníky. Před tím než je výsledný formát vytvořen, jsou rozdělena slova mezerami a znak po znaku se převádí na ASCII znak. Pokud je převod úspěšně dokončen, je vytvořeno jedno slovo ze šablony bez nežádoucích zejména bílých znaků. Takto se převede celá šablona. Pokud se aplikace pokusí převést bílý znak na ASCII znak, je vyhozena a následně zachycena výjimka, která je zpracována. Nově vytvořená šablona je uložena do databáze a pokračuje již dále samotné vyhledávání ceny. Za pomocí knihovny Jsoup



Obrázek 12: Ukázka vygenerovaného grafu pro 2.5 dimenze

je stažena webová stránka a z ní získán div s textem. div je vyhledán pomocí již známého a načteného class nebo id z databáze pro daný e-shop, ke kterému je produkt přiřazen. Vyhledaný text je předán do implementované funkce check, která nalezne veškeré šablony pro daný e-shop. Šablony jsou rozděleny na základě středníku a probíhá obdobné vyhledávání ceny jako u webové aplikace. Vytvořené pole slov z rozdělené šablony je procházeno a vyhledává se každé slovo. Pokud v textu slovo není nalezeno celé vyhledávání pro jednu šablonu je ukončeno a pokračuje se vyhledáváním podle další šablony. Pokud je hledána cena, tak pro maximální spolehlivost je cena vyzkoušena pro převod do datového typu integer nebo float, pokud obsahuje desetinou čárku v ceně. Tak se vyhledají veškeré ceny a dostupnosti podle šablon a nalezené ceny jsou uloženy do kolekce s výsledky. Poté následuje ukládání dat do databáze. V případě, že je cena změněna provede se vložení nového řádku do tabulky epa_priceMovement. Pokud je cena stejná jako při poslední kontrole, provede se pouze aktualizace data kontroly.

7.3 Grafová komponenta

Při vývoji systému ELPOD vznikla potřeba vhodně vizualizovat data v podobě grafů. Jako nejvhodnějším řešením se ukázala JS knihovna Highcharts [28], která umožňuje generování dynamických grafů, které jsou graficky velmi dobře zpracované. Pro usnadnění práce kolegům jsem vytvořil komponentu pro PHP, která umožňuje generování grafů za pomoci knihovny bez znalosti JS. Vytvořená třída pro grafy obsahuje tři základní funkce, které nabízí generování různých grafů. První je funkce singleDimension, vytváří grafy v jedné dimenzi. Takovým představitelem je např. koláčový nebo spojnicový graf s jednou řadou dat. Dále nabízí funkci doubleDimension pro spojnicové, plošné nebo sloupcové grafy obsahující více řad s daty. Třetí funkcí je commonDimension pro tvorbu 2,5 di-

menzionálních grafů 12 . To jsou takové grafy, které obsahují dvě různé osy Y a jednu společnou osu X. Tento graf můžeme využít např. pro přehled teplot a srážek v závislosti na měsících v roce. Společná osa X je časová, tedy měsíce a na jednu osu Y vyneseme teploty a druhou osu vyneseme množství srážek.

Každá funkce pro vytvoření grafu má tři parametry a čtvrtý volitelný, kde se zadává id divu, kam se bude graf na stránce vykreslovat v případě, že strana obsahuje více grafů. První parametr funkce přejímá samotná data pro graf, druhý parametr očekává pole s názvy grafu a popiskami jednotlivých os. Jako třetí parametr se předává pole s nastaveními pro graf. Nastavení umožňuje definovat šířku a výšku výsledného grafu v px, typ vykreslovaného grafu, minimální hodnota na ose Y a nastavení jednotlivých barev jak pro samotné osy, tak pro řady dat. Je možné využít i již předdefinované grafické styly.

Třída obsahuje textovou formu základní kostry grafu. Ta obsahuje značky, které jsou za pomoci funkce `str_replace` v průběhu vytváření grafu postupně nahrazovány textem obsahující nastavení a popiskami grafu. Poslední částí při vytváření grafu je vygenerování datových řad pro Highcharts z předaného pole. Funkce vrátí vytvořený JS kód, který je vložen mezi HTML elementy script. Následně stačí vytvořit div s id, které jsme předali funkci a graf je hotov. Po spuštění internetového prohlížeče JS vykreslí na webové stránce požadovaný graf.

Také pro potřebu generování obrázků z grafu, jsem vytvořil funkce pro její export na stráně našeho serveru. JS vyšle požadavek na server pro export grafu. Data jsou ve formátu SVG, ty jsou na stráně serveru pomocí PHP a třídy `ImageMagick` převedeny do požadovaného formátu a přes http hlavičku je vygenerovaný graf zaslán uživateli ke stažení. Export vytváří grafy ve formátech PNG, JPEG, PDF nebo SVG.

8 Predikce vývoje ceny

Určit přesný vývoj budoucí ceny produktu je víceméně nemožný, ale můžeme se alespoň pokusit o jeho přibližný vývoj. Vývoj ceny je závislý na velkém počtu faktorů. Pro přehled zde uvedu alespoň pár základních:

- cena zlata a dalších důležitých strategických kovů
- politická situace v zemi výrobce a v ČR
- kurz koruny vůči euru
- zvýšení cen pohonných hmot
- mimořádné živelné pohromy
- světové burzy s cennými papíry
- vývoj nových technologií
- ...

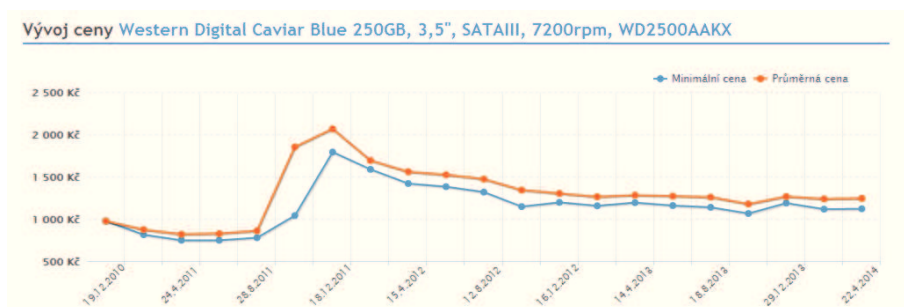
8.1 Zboží ovlivněno externími vlivy

Predikce ceny se vyvíjí podle těchto faktorů, některé lze očekávat ale některé ne. Neočekávanou situací, která výrazně změnila trh, jsou například záplavy v Thajsku v roce 2011. Při záplavách došlo k zasažení výroby pevných disků firem Western Digital a Seagate. Po záplavách došlo k rapidnímu nárůstu cen produktů těchto výrobců. Na to však reagovali zvýšením cen i ostatní firmy jako Samsung a to i když nebyla jejich výroba vůbec zasažena [31, 32]. Graf na obrázku 13 ukazuje růst ceny vlivem záplav.

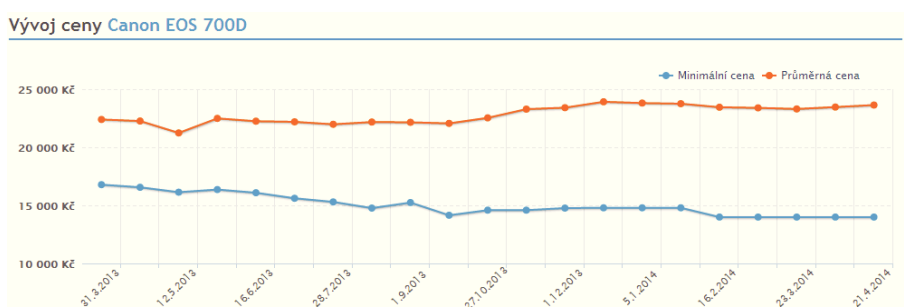
Další výraznou změnou cen zboží, bylo změna kurzu koruny vůči eura. Změnu provedla Česká národní banka k 1.11.2013 a nárůst cen se projevil zejména na elektronice (obrázek s grafem vývoje 14)

8.2 Očekávaný vývoj ceny

Pokud však nebudeme brát na náhodné makro faktory až takový důraz, lze poměrně přesně zajistit budoucí vývoj ceny podle již prodávaných produktů. Pro modelování lze využít již prodávaný starší model produktu, pro kterého chceme určit budoucí vývoj. Po ukázkou využiji prodej mobilního telefonu Samsung Galaxy S. Na obrázku 17 a 18 jsou grafy s vývojem cen modelu S, který byl postupně technologicky inovován. Z grafu je patrné, že přibližně každé tři měsíce poklesne cena přibližně o 500 Kč. Před datem zvýšení prodeje, jako jsou Vánoce, se cena ale mírně zvýší. Přibližně měsíc před nástupem nového modelu začne cena staršího modelu výrazně klesat, ve snaze prodejců vyprodat veškeré skladové zásoby. Dále již klesá cena opět v tříměsíčních intervalech. Graf 19 znázorňuje porovnání cen sledovaných čtyř modelů Samsungu Galaxy, kde lze pozorovat jejich velice podobný vývoj ceny.



Obrázek 13: Graf vývoje ceny pevného disku Western Digital



Obrázek 14: Graf vývoje ceny fotoaparátu Canon

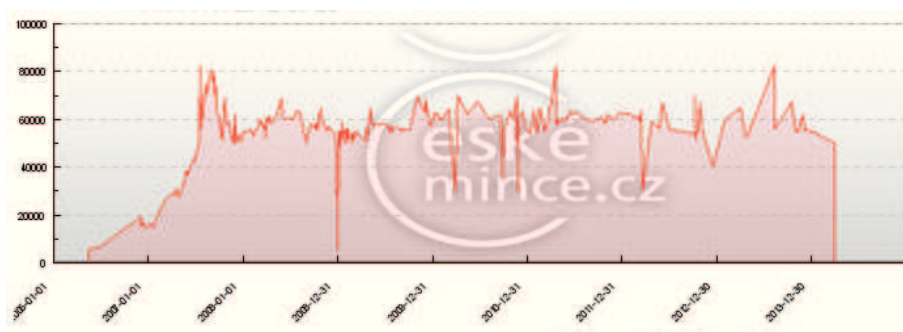
8.3 Zboží s trendem růstu ceny

Obdobný je vývoj u většiny produktů, avšak jiný vývoj je kupříkladu u pamětních mincí. Mince jsou raženy pouze v určitém množství. Po začátku prodeje jejich cena postupně klesá do určité doby. Po vyprodání téměř veškerých zásob se cena mince zvedá, a to z důvodu omezeného množství a její vzácnosti. Na obrázku 15 je graf s vývojem ceny pamětní mince národní kulturní památky papírny ve Velkých Loisnách.

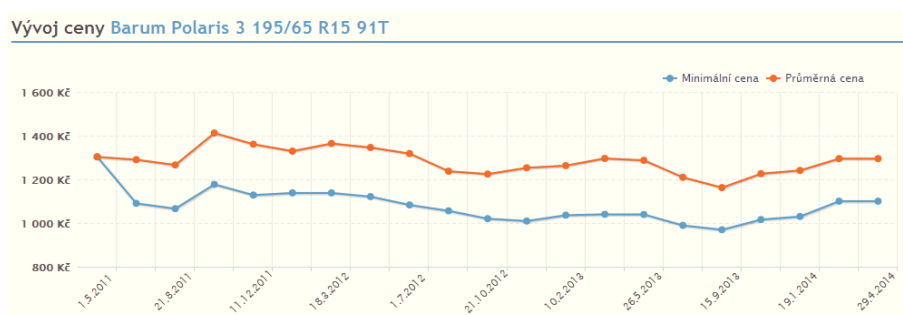
8.4 Sezonní zboží

Dalším typem vývoje je sezonní zboží, jako jsou například zimní bundy nebo zimní pneumatiky. Přibližně měsíc před začátkem období, kdy na autě musí být nazuty zimní pneumatiky, se jejich cena zvyšuje. Tato cena se drží po celou dobu zimy a klesne až počátkem jara. Dále si již drží přibližně stejnou cenu. Pokud se ale na trhu objeví nový model této pneumatiky, cena staršího modelu výrazně poklesne.

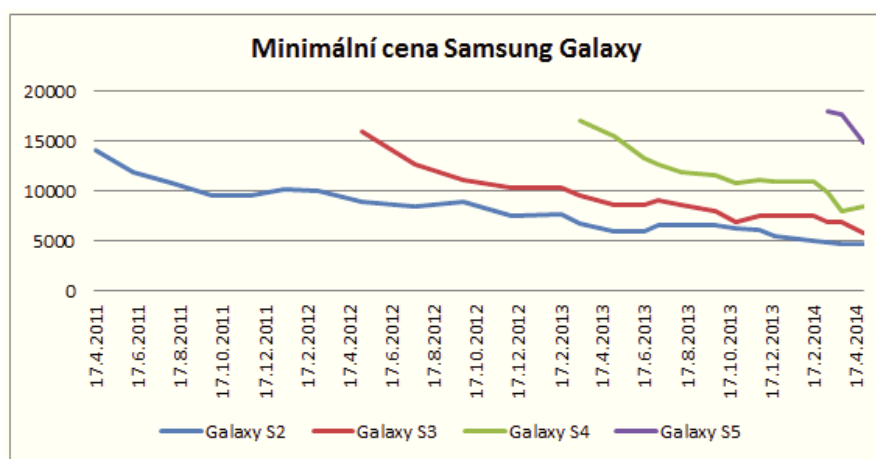
Těchto modelových situací je velká spousta. Pro přibližnou informaci o vývoji ceny jsem tedy těchto získaných informací využil a zakomponoval to do mého nástroje. Pro větší přesnost má uživatel možnost nastavit výchozí cenu produktu, kterou již zná, a od této ceny se bude další vývoj grafu odvíjet.



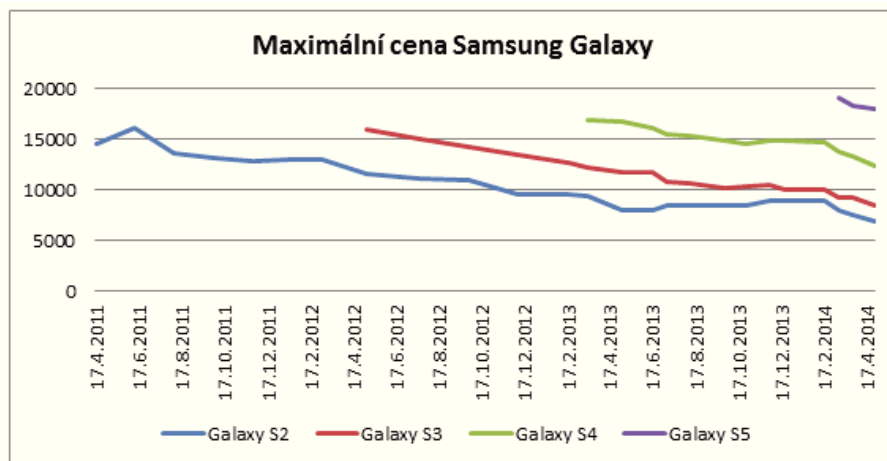
Obrázek 15: Graf vývoje ceny pamětní zlaté mince [35]



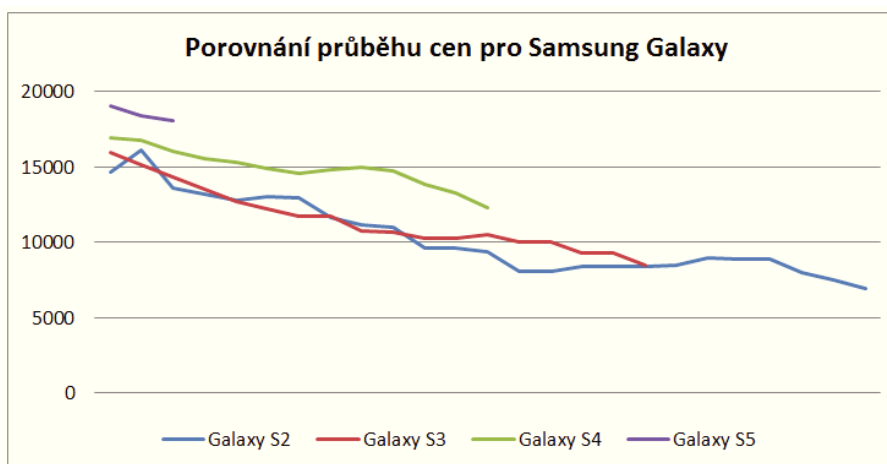
Obrázek 16: Graf vývoje ceny zimních pneumatik [34]



Obrázek 17: Graf vývoje minimální ceny Samsungu řady Galaxy S [33]



Obrázek 18: Graf vývoje maximální ceny Samsungu řady Galaxy S [33]



Obrázek 19: Graf s porovnáním vývoje Samsungu řady Galaxy S [33]

9 Zhodnocení výsledků

9.1 Současný stav a testování

Během implementace i po dokončení byla aplikace testována na velkém množství dat z internetových obchodů. Nelze ale zajistit, že aplikace bude fungovat pro všechny e-shopy, zvláště pro ty, které si prodejci vyvíjejí sami.

Testování probíhalo na lokálním webovém serveru pomocí osobního počítače. Konfigurace počítače byla: CPU: Intel Core 2 DUO - 2,2 GHz, RAM: 4GB, OS: Windows 8 Pro 64-bit a internetové připojení s rychlostí stahování 4 Mbit/s. Pro funkci lokálního serveru byl využit balík aplikací VertrigoServ obsahující webový server Apache, skriptovací jazyk PHP, databázi MySQL s administračním nástrojem PhpMyAdmin. Do budoucna bude nástroj nasazen na produkčním serveru s vyšší výpočetní kapacitou i rychlejším připojením k internetu.

V následujících tabulkách 3 4 je uveden přehled s časy testování aktualizace a import dat s cenami o produktech. Testování jsem několikrát opakoval, abych zjistil přesné časy aktualizace dat. Při použití šesti vláken jsem se dostal na čas 14 s v případě aktualizace dat, která bude v reálném provozu nejčastější akce. Tento čas byl změřen na počtu 110 produktů. I když při aktualizaci dat je pro každý produkt stažena webová stránka, dostal jsem se při aktualizaci dat produktu na čas 0,12 s pro jeden produkt. To je dobrá hodnota, ale pro plné produkční nasazení je to hodnota stále vysoká. Pokud bych chtěl získávat podobné množství dat jako vede Heuréka.cz aktualizace by trvala přibližně 416 hodin (pokud předpokládám, že má Heuréka.cz přibližně 25000 internetových obchodů a každý obchod má cca 500 produktů). V době měření aplikace obsahuje přibližně 60 přesných šablon pro různé internetové obchody a 14 univerzálních vzorů pro obchody založené na systémech jako jsou PrestaShop, OpenCart, Magento nebo české řešení, mezi které patří např. eshoprychle.cz, EasyShop nebo FastCentrik.

Při měření jsem však již narážel na nízký hardwarový výkon osobního počítače a příliš pomalé připojení k internetu, které bylo v době měření maximálně vytíženo. Předpokládám tedy, že po nasazení na server s rychlejším připojením k internetu, budou výsledky aktualizace dat přívětivější. Další možností zrychlení aktualizace je optimalizace algoritmu pro samotné vyhledávání ceny.

9.2 Vize do budoucna

Nástroj bude nasazen v produkčním prostředí pro plné využití uživateli. Bude následovat doladění posledních drobných chyb a v případě velké časové náročnosti systému vyladit samotné algoritmy pro nižší časovou režii.

Počet vláken	Import dat	Průměr
1	54 s	0,49 s
4	17 s	0,15 s
6	12 s	0,11 s

Tabulka 3: Výsledky testování pro import dat

Počet vláken	Aktualizace dat	Průměr
1	55 s	0,5 s
4	20 s	0,18 s
6	14 s	0,12 s

Tabulka 4: Výsledky testování pro aktualizaci dat

Jedná se o poměrně silný nástroj zejména pro začínající podnikatele na internetu. V rámci komplexního systému ELPOD bude opravdu velkým pomocníkem. Systém ELPOD nabízí získávání dat o produktech jak z produktových agregátorů, tak pomocí mého nástroje i ze samotných internetových obchodů. Také nabízí komplexní statistiky o pozicích ve vyhledávačích, zpětných odkazů, ale obsahuje i rádce pro optimalizaci stránek pro vyhledávače nebo zhodnocení podnikatelského záměru.

Jako možné další rozšíření by bylo vhodné získávání základních informací z XML feedů. Pomocí nich by se načetly pouze URL adresy a případné názvy produktů ke spárování. Následně by se vytvořily šablony pro daný e-shop a provedl by se hromadný import produktů, avšak ceny by se již získávaly ze samotných obchodů.

Ze získaných dat můžeme vytvářet podrobné statistiky, ze kterých by bylo možné získávat další informace o vývoji ceny, které by pomohli určovat přesnější predikci vývoje cen pro různé produkty.

10 Závěr

Cílem bakalářské práce byl návrh a implementace nástroje, který bude získávat aktuální a pravdivé informace o cenách a dostupnostech z internetových obchodů, které nejsou zařazeny v produktových agregátorech. Nástroj umožňuje hlídat cenu a dostupnost produktů konkurentů, případně upozorňuje na jejich změnu cen. Rovněž nabízí hlídání aktuálních cen vůči produktovým agregátorům a informuje o případných nekalých praktikách, týkajících se rozdílnosti cen. Získané ceny jsou vizualizovány v grafu pro dokonalý přehled. Součástí nástroje je i funkce zajišťující predikce vývoje ceny z již dostupných dat.

V teoretické části práce jsem se zaměřil na české srovnávače zboží, popsal jejich nabízené funkce a přednosti pro konečné uživatele i prodejce. Porovnal jsem jejich možnosti systémů hodnocení produktů a prodejců, řazení produktů a využití PPC systémů. Dále jsem uvedl zdroje, ze kterých získávají data. Uvedl jsem přehled možnosti získání dat o cenách na internetu a ukázal jsem nejčastěji používané nekalé praktiky v produktových agregátorech. V praktické části byla provedena datová a funkční analýza systému. Popsal jsem zde požadavky na vyvíjený systém a jeho funkce. V části návrhu implementace jsem vybral vhodné a moderní technologie.

Při implementaci jsem vyvinul grafovou komponentu, která ulehčila práci se zobrazením dat mým kolegům, spolupracujícím na systému ELPOD.

Výsledkem je funkční nástroj vhodný pro internetové obchodníky. Oproti jiným obdobným nabízeným službám na tuzemském internetu nabízí získání reálných dat z velkého množství internetových obchodů.

11 Reference

- [1] ČR je zemí e-shopům zaslíbená *Shoptet* [online]. 2012 [cit. 2014-02-15]. Dostupné z: <http://blog.shoptet.cz/cr-je-zemi-e-shopum-zaslibena/>
- [2] Měsíční zpráva únor 2014. *NetMonitor* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://www.netmonitor.cz/unor-2014>
- [3] Historie a vývoj e-shopů v podmínkách českého trhu. *Finarea* [online]. 2013 [cit. 2014-02-15]. Dostupné z: <http://www.finarea.cz/ekonomicke-jevy/123-historie-a-vyvoj-e-shopu-v-podminkach-ceskeho-trhu/>
- [4] Grafy a statistická data. *APEK* [online]. 2013 [cit. 2014-02-15]. Dostupné z: <http://www.apek.cz/grafy-a-statisticka-data/>
- [5] Historie elektronických obchodů. *Marketingové noviny* [online]. 2006 [cit. 2014-02-15]. Dostupné z: http://www.marketingovenoviny.cz/marketing_4391/
- [6] RIZIKOVÉ NÁKUPY PŘES INTERNET. ČOI [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://www.coi.cz/rizikove-nakupy-pres-internet-nc1012/>
- [7] Zlatý slitek Perth Mint 1 oz *Heuréka.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://investicni-zlato.heureka.cz/zlaty-slitek-perth-mint-1-oz/>
- [8] Ověřeno zákazníky. *Heuréka.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://sluzby.heureka.cz/napoveda/certifikat-spokojenosti/>
- [9] Jak funguje řazení obchodů a PPC systém? *Heuréka.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://sluzby.heureka.cz/napoveda/ppc/>
- [10] Měření konverzí. *Heuréka.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://sluzby.heureka.cz/obchody/mereni-konverzi/>
- [11] Doplnkový XML soubor pro dostupnost zboží. *Heuréka.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://sluzby.heureka.cz/napoveda/dostupnostni-feed/>
- [12] Specifikace XML souboru. *Heuréka.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://sluzby.heureka.cz/napoveda/xml-feed/>
- [13] FAQ (často kladené dotazy). *Zboží.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://napoveda.seznam.cz/cz/zbozi/faq/>
- [14] Uživatelské hodnocení obchodů (recenze) *Zboží.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://napoveda.seznam.cz/cz/zbozi/napoveda-pro-internetove-obchody/hodnoceni-obchodu/>
- [15] Specifikace XML pro internetové obchody. *Zboží.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://napoveda.seznam.cz/cz/zbozi/napoveda-pro-internetove-obchody/specifikace-xml/>

-
- [16] Obchody v okolí. *Google Merchant Center* [online]. 2014 [cit. 2014-02-15]. Dostupné z: https://support.google.com/merchants/answer/1101441?hl=cs&ref_topic=3404821
- [17] Začínáme s kampaněmi v Nákupch a účtem AdWords *Google Merchant Center* [online]. 2014 [cit. 2014-02-15]. Dostupné z: https://support.google.com/merchants/answer/2660968?hl=cs&ref_topic=2660962&rd=1
- [18] Přehled zdrojů dat. *Google Merchant Center* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <https://support.google.com/merchants/answer/188478>
- [19] Specifikace zdroje produktů. *Google Merchant Center* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <https://support.google.com/merchants/answer/188494?hl=cs&topic=2473824&ctx=topic#CZ>
- [20] Registrace. *Srovnáváme.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://www.srovname.cz/muj-obchod/registrace?step=krok1>
- [21] Popis XML. *Srovnáváme.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://www.srovname.cz/muj-obchod/xml-feed>
- [22] Měření konverzí. *Srovnáváme.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://www.srovname.cz/muj-obchod/mereni-konverzi>
- [23] Ceník prokliků. *Srovnáváme.cz* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://www.srovname.cz/muj-obchod/cenik>
- [24] About rich snippets and structured data. *Webmaster Tools Help* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <https://support.google.com/webmasters/answer/99170?hl=en>
- [25] PHP Simple HTML DOM Parser. *Source Forge.net* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://simplehtmldom.sourceforge.net/>
- [26] Jsoup: Java HTML Parser. *JSOUP* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://jsoup.org/>
- [27] Nette Framework. *Nette* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://nette.org/cs/>
- [28] Highcharts and Highstock documentation. *Highcharts JS* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://www.highcharts.com/docs>
- [29] PHP: Hypertext Preprocessor *PHP* [online]. 2014 [cit. 2014-02-15]. Dostupné z: <http://www.php.net/>
- [30] Použitelné znakové entity. *Jak psát web* [online]. 2014 [cit. 2014-04-10]. Dostupné z: <http://www.jakpsatweb.cz/html/entity-vsechny.html>
- [31] Zdražení pevných disků kvůli záplavám v Thajsku pocítí i výrobci základních desek. *diit.cz* [online]. 2011 [cit. 2014-04-10]. Dostupné z: <http://diit.cz/clanek/zdrazeni-pevnych-disku-kvuli-zaplavam-v-thajsku-pociti-i-vyrobci-zakladnich-desek>

- [32] Disky po zdražení: co se změnilo a co je nejvýhodnější. *EXTRAHARDWARE.CZ* [online]. 2011 [cit. 2014-04-10]. Dostupné z: <http://www.cnews.cz/clanky/disky-po-zdrazeni-co-se-zmenilo-co-je-nejvyhodnejsi>
- [33] Samsung G900 GALAXY S5. *Heuréka.cz* [online]. 2014 [cit. 2014-04-10]. Dostupné z: <http://mobilni-telefony.heureka.cz/samsung-g900-galaxy-s5/>
- [34] Barum Polaris 3 195/65 R15 91T. *Heuréka.cz* [online]. 2014 [cit. 2014-04-10]. Dostupné z: <http://pneumatiky.heureka.cz/barum-polaris-3-195-65-r15-91t/>
- [35] Národní kulturní památka papírna Velké Losiny. *České-mince.cz* [online]. 2014 [cit. 2014-04-10]. Dostupné z: http://www.ceske-mince.cz/coin_gold_proof/2006/papirna-velke-losiny-zlata-mince-proof/

A Obsah CD

bakalarska-prace.pdf	- text bakalářské práce
src/	- zdrojové kódy
java/	- implementace parseru
admin/	- implementace administrace webové aplikace
front/	- implementace uživatelské webové aplikace
tools/	- pomocné nástroje
data/	- databázové skripty